DOCUMENT RESUME

ED 067 379 24 SP 005 871

AUTHOR Brophy, Jere E.
TITLE A Study to Determine if Teachers Communicated Differential Performance Expectations to Students. Final Report.
INSTITUTION Texas Univ., Austin. Research and Development Center for Teacher Education.
SPONS AGENCY National Center for Educational Research and Development (DHEW/OE), Washington, D.C.
BUREAU NO BR-1-F-012
PUB DATE Aug 72
CONTRACT OEC-6-71-0477-(509)
NOTE 76p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Communications; *Educational Research; *Performance Criteria; *Performance Factors; *Student Teacher Relationship; Teacher Response
IDENTIFIERS Dyadic Observation System

ABSTRACT

This study was planned to replicate and extend at a fifth grade level an earlier study by Brophy and Good at the first grade level. One purpose was to test the hypothesis that classes taught by teachers who showed evidence of expectation effects would show polarization over time, with differences between high and low expectation students gradually becoming increased. A second purpose was to investigate the form in which expectation effects would be manifested at the fifth grade. Subjects included five fifth grade teachers and their respective students. The research design and methods involved systematic naturalistic observation rather than experimental manipulation or treatment. Teachers' naturalistically formed expectations were determined, and then student-teacher interaction was observed with a version of the Dyadic Observation System to see if teachers show favoritism toward high expectation students or inappropriate treatment of low expectation students. Results suggest that student ability level does not affect the stability of classroom interaction measures, and that correlations between measures taken in different subject matter classes taught by the same teacher tend to be only very slightly higher than correlations taken in classes involving the same subject matter taught by two different teachers. A 19-item bibliography and a teachers' ranking form are included. (Author/MJM)

FINAL REPORT
Project No. 1F012
Contract No. OEC-6-71-0477-(509)

A STUDY TO DETERMINE IF TEACHERS COMMUNICATED DIFFERENTIAL PERFORMANCE EXPECTATIONS TO STUDENTS

Jere E. Brophy
Department of Educational Psychology
The University of Texas at Austin
Austin, Texas 78712

August 1972

U. S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

TABLE OF CONTENTS

LIST OF TABLES

A C K N O W L E D G E M E N T S

The author wishes to extend recognition and appreciation to the following individuals for their assistance in carrying out this project: Dr. Thomas Good, who helped plan the original design; Dr. Carolyn Evertson, who coordinated data collection and preparation; Susan Florence, Dr. Teresa Harris, Karen Mays, Mark Mays, Piara Pannu, Brian Peck, and Kathleen Senior, who served as classroom coders; Shyam Ahuja, Marilyn Arnold, Maria Bucynski, Dr. Teresa Harris, Carol King, Mark Mays, Piara Pannu, and Michael Weissberg, who helped prepare data for analyses; and Brian Peck and John Sheffield, who performed the statistical computations.

Data preparation and analysis support was provided by the Research and Development Center for Teacher Education at the University of Texas at Austin, Oliver H. Bown and Robert F. Peck, co-directors.

SUMMARY

This study attempted to replicate at the fifth grade level the findings of Brophy and Good (1970a) regarding teachers' communication of differential performance expectations to individual students at the first grade level. It also attempted to test the polarization hypothesis from the Brophy and Good (1970a) model for teacher expectation effects, which suggests that differential teacher treatment of high and low expectation students will cause these two groups of students to become increasingly more different from each other in their classroom behavior and achievement levels as the school year progresses. A secondary purpose of the study was to test the hypothesis that teacher expectation effects would be mediated more through quantitative measures of teacher-student interaction at the fifth grade level as compared to the first grade level.

The findings regarding replication of the Brophy and Good (1970a) data were negative. There were few significant differences between high and low expectation students on 49 measures of teacher-student interaction, and the significant differences which did appear included none of the important indicators of communication of communication of differential performance expectation that had appeared in the Brophy and Good (1970a) study. Thus, the Brophy and Good (1970a) results were not replicated. Because of this, the polarization hypothesis could not be tested, since it assumes an already existing favoritism of high expectation students over low expectation students, and such favoritism did not exist in this sample of teachers as a group. Two of the five teachers involved did show favoritism of high over low expectation students, but analyses of their data provided little support for the polarization hypothesis. Thus, this hypothesis remains essentially untested in a sample of teachers large enough to allow a formal statistical test of the significance of changes in teacher-student interaction patterns over time, although the nonstatistical case studies of teachers who did show evidence of expectation effects were not encouraging with regard to the polarization hypothesis.

The data from this study also allowed investigation of stability of teacher-student interaction patterns within the same classrooms over time, and, because the same students were observed in two or more classrooms, investigation of stability across classrooms within the same time period. These stability coefficients indicated a moderate degree of stability in the interaction measures (about two-thirds of the coefficients, including all which reached the .05 level of statistical significance, were positive). However, as is typically the case with classroom data based on coding of discrete interactions, they were less impressive than stability coefficients coming from high inference ratings or other measures based on observers' global judgments. Stability coefficients within the same classroom across time periods were generally higher than stability coefficients across different classrooms within the same time periods. Student ability level had little effect on the stability coefficients, and stability coefficients

were only very slightly higher when they came from classes involving different subject matter taught by the same teacher rather than classes involving the same subject matter taught by different teachers. In general, stability coefficients were positively correlated with the frequency of observation of classroom interaction variables, suggesting that they would have been higher if more data had been collected or if cut off points had been established to eliminate correlations based on data from only a few students.

I N T R O D U C T I O N

Publication of Rosenthal and Jacobson's (1968) Pygmalion in the Classroom generated interest and controversy which has continued since, both in the lay public and among educational researchers, concerning the hypothesis that teacher expectations act as self-fulfilling prophecies. In their experiment, Rosenthal and Jacobson identified a few children in each of a group of elementary school teachers' classrooms as being "late bloomers," who could be expected to show unusually large achievement gains in the coming academic year. Although the "bloomers" had been selected from the class rosters at random, teachers were given the impression that they had been identified by a test given to the children. The test was actually a general abilities test, but the teachers did not know this.

The treatment or intervention in this study consisted of a single interview in which these "late bloomers" were identified to the teachers. However, Rosenthal and Jacobson presented achievement test data suggesting that the "late bloomers" had indeed made greater than expected gains during the academic year than their classmates made. They attributed their findings to the "Pygmalion" or "self-fulfilling prophecies" effect of teacher expectations. By raising the teacher expectations regarding the "late bloomers," they reasoned, they had also changed teacher behavior in some way which led to the teachers' actually producing more achievement in the "late bloomers" than they did in their classmates.

Publication of these results set off a flurry of debate, with some observers accepting the data enthusiastically, others suggesting that the findings held only for grades one and two but not for grades three through six, and still others suggesting that there were so many methodological problems with the study that the data could not be accepted at all (For a summary of this debate, including favorable and unfavorable reviews, charges by critics, and the responses by Rosenthal and Jacobson, see Elashoff and Snow, 1971).

Initial reactions to the study were prematurely and often overly enthusiastic. Sometimes writers even seemed to suggest that any teacher expectation would somehow automatically or magically become self-fulfilling. Later, however, publication of several negative reviews and several failures to replicate the findings (summarized in Elashoff and Snow, 1971) caused this early enthusiasm to be replaced with a much more negative view. As a result, it was commonly stated that the hypothesis that teachers' expectations could act as self-fulfilling prophecies had been disproven, or at least that there was no solid evidence to support it. This negative view still persists in some quarters, partially because the debate over Rosenthal and Jacobson's original study has kept attention focused on it, and has diverted attention from several other studies done in the meantime which provide solid evidence to show that teachers' expectations can and sometimes do function as self-fulfilling prophecies, although they do not always do so (Brophy and Good, 1972).

To place the present research in context, it will be helpful to introduce two major dimensions upon which studies in this area can be categorized: experimental inducement of teacher expectations vs. the use of teachers' naturalistically formed expectations, and linking teachers' expectations to student gain data vs. linking teachers' expectations to teacher-student interaction measures. These two dimensions are independent of each other. Studies involving experimentally induced teacher expectations could attempt to link these expectations to either student gain data or to teacher-student interaction data. Similarly, studies of teachers' naturalistically formed expectations could attempt to link these expectations to either student gain data or to student-teacher interaction data. In fact, all four types of studies have been done.

The present study, and all of the others in which the author has been involved, used the teachers' naturalistically formed expectations. This was done for two reasons. First, this method is simpler, more direct, and more reliable. An investigator can elicit a teacher's expectations regarding his students' performance simply in this case, for example, by asking him to rank the students according to expected achievement. Thus teacher expectations are measured directly, and the experimenter can be sure that the expectations he assumes the teacher to hold are the expectations that the teacher does in fact hold (unless the teacher, for whatever reason, is not truthful in reporting his expectations). In contrast, experimenters attempting to induce expectations in teachers through some kind of manipulation or treatment cannot be sure that each teacher did in fact acquire the expectations that the experimenter wanted him to acquire. Thus if negative results are obtained, the experimenter does not know whether it was because the teachers did not acquire the desired expectations or because the teachers' expectations did not influence their interactions with students.

A second reason for using teachers' naturalistically formed expectations is that the data from this type of study are more directly generalizable to the average or typical classroom than are data from an experimental or manipulative study. The average teacher does not have psychologists or other investigators inducing his expectations regarding students, but he does naturalistically form his own expectations about students in his everyday interactions with them (as well as through reading school records, talking to other teachers, etc.).

Thus the author favors the use of teachers' naturalistically formed expectations over attempts to induce expectations experimentally, because the former method is simpler and more direct and because it allows for more direct generalizations to typical classrooms. For a more detailed discussion and review of literature relevant to these points, and for data showing that studies using teachers' naturalistically formed expectations are much more likely to show positive results than studies involving experimentally induced expectations (primarily because in many of the experimental studies the teachers simply did not acquire the expectations that the experimenters attempted to induce in them), see Brophy and Good, 1972.

Although the distinction between studies using experimentally induced expectations and studies using teachers' naturalistically formed expectations is relevant to the present research for the reasons stated, the distinction between studies linking teacher expectations to student gain data as opposed to studies linking teacher expectations to teacher-student interaction data is even more central to the basic purpose of this study. Consequently it will be discussed more fully, and certain relevant studies will be reviewed.

Studies linking teacher expectations to measures of student gain are both interesting and important, because they show that teacher expectations can influence the amount that a student learns. Whether or not one accepts the Rosenthal and Jacobson data, there have been other studies done since which show convincingly that teachers sometimes produce greater gain in students when they have high expectations for them and lesser gains when they have lower expectations for them (Palardy, 1969; Doyle, Hancock, and Kifer, 1971; Tuckman and Bierman, 1971). The word "sometimes" was included in the previous sentence because it has now been established that although teachers' expectations can act as self-fulfilling prophecies, they do not always or necessarily do so (Brophy and Good, 1973).

A Model for Research on Teachers' Expectations

Thus studies linking teachers' expectations to the degree of learning that students show in the teachers' classrooms are important for documenting the fact that teachers' expectations can influence student learning. However, although such studies document the fact that teacher expectations can act as self-fulfilling prophecies, they reveal nothing about how the process works. Studies linking teacher expectations to teacher-student interactions are required to reveal the mechanisms involved when teachers' expectations act as self-fulfilling prophecies (hereafter, the phrase "the self-fulfilling prophecy effect of teachers' expectations" will be replaced by the shorter phrase "teacher expectation effects").

The present study, like several others the author has been involved in (with Dr. Thomas L. Good and with several doctoral students), was of the latter type. It attempted to relate teachers' expectations for different students to differential teacher treatment of those students, and thereby to deepen our understanding of how teachers communicate differential expectations to students so that the performance of high expectation students is maximized while that of low expectation students is depressed. Based on the studies referenced above (and on others reviewed in Brophy and Good, 1973), this research assumed the existence of teacher expectation effects as an established fact. Consequently, its focus was not on documenting such effects but on identifying the mechanisms that produce them.

The basis for the study was the theoretical model for teacher expectation effects presented by Brophy and Good (1970a). Reasoning that the mechanisms underlying teacher expectation effects must lie in obser-

vable and measurable behavior rather than in some mysterious process akin to ESP, Brophy and Good (1970a) developed the following model as a tentative explanation for teacher expectation effects and as a guide to systematic research in the area:

1. Early in the year the teacher forms differential expectations for the academic performance of different children in the classroom;
2. The teacher then begins to treat the different children in accordance with his expectations for their performance;
3. The children will then begin to react to their teacher differentially because they are being treated differently, and this reactive behavior will tend to complement and reinforce the teacher's expectations;
4. The cumulative effects of this sequence of events will be seen in the achievement test scores at the end of the school year, which will provide objective evidence that teachers' expectations function as self-fulfilling prophecies.

In their original study Brophy and Good (1970a) undertook to establish whether or not there were data to support Step 2 of the model. Documentation for Step 1 was unnecessary, since teachers can readily express differential performance expectations for their students, even as early as the first day of school (Brophy and Good, 1973). Among other things, this study required the development of a new classroom interaction observation system, the Brophy-Good Dyadic Interaction System (Brophy and Good, 1970b). Previously developed systems had focussed almost exclusively upon teacher behavior and had treated the class as an undifferentiated unit, so that information on the teachers' interactions with each different student in the class was not preserved in the coding. A key feature of the Dyadic System is that it uses the individual student rather than the class as the unit of analysis, so that teacher-student interactions are separately coded and recorded for each separate student in the class. In addition, the coding system allows for the preservation of the sequence in which interaction events occur, so that a given interaction can be classified as teacher-initiated or student-initiated and so that the details of the sequential order of chains of interaction can be preserved. This feature makes possible certain inferences about cause and effect relationships between some of the variables, in addition to yielding information about correlation among the variables that other coding systems also yield.

The original experiment (Brophy and Good, 1970a) yielded clear cut evidence of teacher expectation effects. In this research, four first grade teachers were asked to rank their students according to the level of achievement they expected from them. Using these rankings, three high and three low boys, and three high and three low girls, were identified for observation in each class. Teacher-child interaction was then observed for 10 hours in each classroom, using the Dyadic System. Many differences in interaction patterns between the high and the low group were observed, although a majority of these were attributed to differences in the behavior of the children themselves rather than to a tendency of the teacher to discriminate in favor of highs at the expense

of lows. Thus, for example, highs had a greater percentage of correct answers, fewer errors per reading turn, higher frequencies of hand raising, more student-initiated individual interactions with the teachers, and fewer interactions involving teacher correction or criticism of misbehavior. These group differences are wholly or partially attributable to differences in the classroom behavior of the children in the high and low groups, and thus cannot be taken as evidence of teacher behavior involved in communicating performance expectations to students.

However, evidence of the latter sort was observed on several measures of differential teacher treatment of the two groups of children in parallel situations. For example, both in general class activities and in reading groups, the teachers more frequently stayed with highs after they failed to answer an initial question correctly, extending the interaction and providing a second response opportunity for these students by repeating the question, giving a clue, or asking another question. In contrast, they were much less likely to stay with lows in these situations, and much more likely to end the interaction by giving the answer or calling on someone else.

In addition to these differences in persistence in seeking responses from the students, several differences in teacher feedback reactions to students were noted. One difference concerned teacher failure to give any feedback at all following a student response. Teachers failed to give feedback to highs in only 3% of their response opportunities, but they failed to give feedback to lows in 15% of their response opportunities. Furthermore, highs were more likely to be praised when they answered correctly and less likely to be criticized when they answered incorrectly or failed to respond, even though they had many more correct answers and many fewer failures than the lows.

In summary, the teachers in this study treated the highs more appropriately, working to obtain good responses from them and reinforcing such responses when they succeeded in obtaining them. In contrast, they gave up easily rather than persist in trying to obtain responses from the lows, and even when they did obtain good responses, they often failed to reinforce them appropriately. Thus they were slower to praise and quicker to criticize the students who most needed patience and encouragement.

These data were taken as evidence of teacher expectation effects, because if continued over time they would maximize the performance of the highs and depress the performance of the lows. The differences in teacher persistence in working with the two groups of students would provide the highs with greater opportunity to learn than the lows, and the differences in responding to and in praising and criticizing student answers would tend to enhance the motivation of highs but depress the motivation of lows and perhaps even alienate them from the teachers. Thus this study succeeded in identifying some of the ways in which teachers communicate differential performance expectations to students, and, by inference, some of the mechanisms explaining teacher expectation effects. Other studies in this vein (linking teachers' expectations to teacher-student interaction data) will be reviewed below.

Related Studies

Beez (1968) studied teachers working in tutorial situations with Headstart students who had been randomly labeled as either high or low ability (thus the teacher expectations were experimentally induced in this study). The teachers who tutored children whom they thought to be high ability students taught more than the teachers working with students that they thought to be of low ability, even though the students had been assigned to the two groups at random. Furthermore, when questioned after the task, only three per cent of the tutors working with "high ability" students thought that the task was too hard for the students, while 63% of the tutors working with "low ability" students thought that the task was too hard. Student learning was also assessed, and it was found that "high ability" students scored higher than the "low ability" students. The amount learned was directly related to the amount that teachers attempted to teach. Thus this study illustrates that one mechanism explaining teacher expectation effects is the relationship between expectation and student opportunity to learn. Teachers apparently attempt to teach more to students whom they expect to learn more and attempt to teach less to students whom they expect to learn less.

Rothbart, Dalfen, and Barrett (1971) studied the behavior of student teachers working with groups of four different ninth graders. Two of each group of four students had been labeled as "lacking in intellectual potential," while the other two had been described as having "considerably greater academic ability." Although the teachers directed equal amounts of reinforcement towards the two groups of students, they were more attentive towards the "brighter" ones. They also rated the "brighter" students as more intelligent, as having greater potential for future success, and as having less need for approval. Thus this study illustrated that teachers may be more attentive towards students whom they perceive as brighter or as having greater potential. This finding is consistent with the finding regarding teacher failure to provide feedback in the Brophy and Good (1970a) study, in that the teachers' more frequent failure to provide feedback to lows suggests that they were paying less careful attention to the responses of lows.

Rubovits and Maehr (1971) studied undergraduate volunteer teachers working with groups of fourth, sixth, and seventh graders in microteaching situations. The students had been labeled as gifted or non-gifted. In these situations, the teachers requested more statements, initiated more interactions, and directed more praise toward the "gifted" students than towards the others. The data on praise are consistent with those reported by Brophy and Good (1970a). The data on requesting statements and initiating interactions introduce a new element: differential quantity of interactions, with teachers perhaps disposed to interact more frequently with high expectation than with low expectation students.

Medinnus and Unruh (1971) observed Headstart teachers working with students enrolled in their own classrooms. The investigators matched pairs of boys whose IQ's were in the 95 - 105 range, but then identified

one of each pair to his teacher as a "high ability" child with an IQ "above 105," and designated the second as a "low ability" boy with an IQ "below 95." Teachers were then observed as they worked with the students individually in teaching a block sorting task. Consistent with findings already reported above, the teachers in this study directed more praise and less criticism toward the "high ability" students.

Meichenbaum, Bowers, and Ross (1969) found that teachers gave more positive and less negative attention to students identified as "late bloomers" than to matched controls, and that the "late bloomers" later outperformed controls on objective tests.

Except for the Brophy and Good (1970a) study, the preceding studies all involved experimental inducement of teacher expectations. In addition to these experimental studies, several studies of the relationship between teacher expectations and teacher-student interaction have been conducted using teachers' naturalistically formed expectations.

Rowe (1969) asked elementary school science teachers to indicate who the top and bottom five students were in the classes. She then observed the teachers' wait time during question and answer sessions. The teachers were timed to see how long they would wait for a response after questioning a student. It was found that the teachers waited twice as long for a response from the top group than they waited for a response from the bottom group. Thus the students least able to respond had to do so more quickly or lose their chance, while the students most able to respond were given more time to answer. Rowe also found that the bottom group received both more criticism and more praise from the teachers. However, she noted that praise directed toward the bottom group was less specific and generally less appropriate than praise given to the top students. Top students were praised for correct responses, while the bottom students were sometimes praised for incorrect responses. The latter finding again suggests that perhaps teachers are not paying as close attention to the responses of bottom students as they are to those of top students (although other interpretations are possible).

When Rowe persuaded the teachers to increase their wait time, she found that the length and quality of student responses increased, and that the frequency of unsolicited student suggestions and comments increased, also. In addition to these findings for the class as a whole, she noted that the distribution of student contributions to discussions began to be spread more widely and evenly across the class. Students in the bottom group who had in the past contributed relatively little to class discussions began to speak up more often, sometimes enough to come to the teachers' attention and begin to change their expectations.

Rist (1970) reported a case study in which he observed teachers' interactions with the same group of children from the beginning of kindergarten through the end of second grade. Although no formal data are included, Rist described several gripping incidents in which teacher expectations were quite directly communicated to the children. He also reported that the teachers interacted more often and more positively with the high expectation students.

The Present Study

Although the list is probably incomplete and some inconsistencies need to be traced down before they can be fully understood, the studies linking teacher expectations to differential patterns of interaction with different students, when taken together, provide abundant evidence to support Step 2 of Brophy and Good's (1970a) model for teacher expectation effects. That is, there is abundant evidence that when teachers hold different expectations for different students, they tend to treat such students differently in accordance with those expectations. Some of the ways that already have been identified were listed above. Thus Step 2 of the model can be considered to be confirmed at this point.

Evidence is lacking to date regarding the third step of the model, however. This step predicts that once the teachers begin to treat children differentially, the children will begin to respond differentially. A direct corollary of this is the prediction that the class will become more polarized over time, such that a greater difference will exist between the highs and the lows later in the semester or the school year than existed earlier in the semester or the school year. The rationale here is that positive teacher treatment toward high expectation students will maximize their learning as well as their interest in and motivation regarding school activities, while negative teacher treatment of low expectation students will minimize their learning and probably also minimize their interest and alienate them from the teacher and from classroom learning experiences. Over time they should begin to approach the teachers less frequently, persist in their work less diligently, and otherwise show signs of deterioration in performance and motivation.

Meanwhile, the high expectation students who are receiving both good teaching and good positive reinforcement from the teachers should respond with enthusiasm and persistent learning efforts. Such differences would in turn increase and reinforce the teachers' differential expectations, which again in turn would tend to increase and reinforce the teachers' tendencies to discriminate between the two groups. Once such a vicious circle is set in motion, the hypothesis that the class would become more polarized over time emerges.

One purpose of the present study was to test this "polarization" hypothesis by seeing if high and low groups would become more different from each other as the semester progressed in classrooms where evidence of teacher expectation effects was detected. This polarization hypothesis has not been effectively tested to date. One attempt to test it in a study of 9 first grade teachers (Brophy and Good, 1973) produced ambiguous and uninterpretable results. Only three of the 9 teachers involved showed evidence of teacher expectation effects (favoritism towards highs and discrimination against lows). Among these three, one showed evidence of polarization within her classroom, one showed no trend, and one showed the opposite trend. Thus the polarization hypothesis, Step 3 of the model, remained untested prior to the present research.

Another purpose of the study was simply to replicate other work that had been done at the first grade level. Two reasons indicated the need for such replication. First, many of the studies cited previously were done at the preschool or first grade levels. Also, as noted previously, in the original Rosenthal and Jacobson study, evidence for teacher expectation effects was confined largely to the first two grades, especially to the first grade. Differences between "late bloomers" and their classmates in grades 3 to 6 in that study were very minor. Teacher expectation effects involving students at higher grade levels have been shown in a few experimental studies, but at the time this study was undertaken, no evidence had been published to document teacher expectation effects in naturalistic studies involving teachers working with their own students under normal conditions other than in the first few grades of elementary school.

Some have argued that because early elementary students have not yet established a reliable "track record" which will allow their teachers to make extremely accurate predictions about their performance in a given school year, the potential for teacher expectation effects may be higher in these early grades. Some have even suggested that teacher expectation effects will be minimal or even nonexistent in higher grades, once the students have "established themselves" relative to their classmates. Hence the need for a study to establish whether or not the indicators of teacher expectation effects observed in Brophy and Good's (1970a) first grade study would be replicated in a study at the fifth grade level.

Another reason for replicating at this level is the suggestion that teacher expectation effects might be present but might be mediated through different mechanisms by teachers working with older students than by teachers working with younger students (Brophy and Good, 1973). If one makes a distinction between the quality (appropriate vs. inappropriate) and the quantity (frequent vs. infrequent) of interactions that a teacher and student might share, there is some evidence to suggest that teacher expectation effects will be mediated primarily through qualitative measures of teacher-student interaction in the early grades, while expectation effects in later grades will be mediated primarily through quantitative measures (Brophy and Good, 1972).

Reading groups and other mechanisms for instruction used in the early elementary grades tend to equalize the frequency or quantity of contacts that teachers have with the different students in their classroom at this level. Thus, for example, Brophy and Good (1970a), despite finding a number of qualitative indicators of differential treatment of high and low expectation students in first grade classrooms, did not find any difference in the total number of contacts between teachers and students in the two groups. The sheer numbers of contacts that each group had with their teachers were usually about even. However, a greater proportion of the contacts involving high expectation students were initiated by the students themselves, a greater proportion dealt with academic or work-related matters, and a lesser proportion dealt with disciplinary matters. These data and data from other

studies suggest that early elementary grade teachers usually do not interact more frequently with high expectation students than they do with low expectation students. However, the quality of the interactions that they do have with high expectation students tends to be higher (more appropriate, more positive, more facilitating) than the quality of the interactions that they have with low expectation students.

In addition, other data (Horn, 1914; Jackson and Lahaderne, 1967; Brophy and Good, 1973) suggest that with each advancing grade level the high achieving students begin to dominate more and more of the class discussions and question and answer sessions, while the low achieving students become increasingly passive or at least non-participatory. Differences are especially notable in public interactions (discussions or question and answer sessions). They are less pronounced in private interactions; teachers are more equal in going from student to student to check work at the individual students' desks. In any case, it can be predicted from several studies that if teacher expectation effects are observed in the later elementary or secondary grades they are more likely to show up as quantitative differences (teachers having many more interactions with highs than with lows) than as qualitative differences.

This could have been studied in the extreme case if the present research had been conducted in late secondary or even college classrooms. However, the differences between these situations and the first grade are so great as to introduce an uncomfortably large number of alternative explanations for any discrepant results that might occur. Thus, the fifth grade was selected as an intermediate point that should be far enough removed from the first grade for the hypothesized qualitative to quantitative shift to occur if it exists, and yet close enough on several other relevant variables (both the students and the teachers are in elementary school classrooms; each class has one homeroom teacher with whom they are primarily identified and with whom they spend most of their day; teachers and students still have a relatively personal relationship compared to the relatively impersonal relationship existing at the high school and college levels), so that meaningful comparisons between the first grade and fifth grade data could be made.

Summary

This study was planned to replicate and extend at a fifth grade setting an earlier study by Brophy and Good (1970a) at the first grade level. One purpose was to test the hypothesis that classes taught by teachers who showed evidence of expectation effects would show polarization over time, with differences between high and low expectation students gradually becoming increased. A second purpose was to investigate the form in which expectation effects would be manifested at the fifth grade, because some evidence suggested that they would be more likely to show up in quantitative rather than qualitative measures of teacher-student interaction.

METHOD

The research design and methods used in this study were variations of the same basic method used by the author in a series of studies done in collaboration with Dr. Good and other colleagues (Brophy and Good, 1973). The method involves systematic naturalistic observation rather than experimental manipulation or treatment. Teachers' naturalistically formed expectations are determined, and then teacher-student interaction is observed with a version of the Dyadic Observation System to see if teachers show favoritism towards high expectation students or inappropriate or biased treatment of low expectation students.

Subjects

Subjects for the research were five fifth grade teachers and their respective students. The students were overwhelmingly white, although those in one school were predominantly middle class while those in the other school were predominantly upper lower class. Schools were deliberately chosen to reflect these two different social class levels to see if the social class of the student populations involved would have any effect on the data. The study was limited to schools serving predominantly white student populations to avoid confounding possible racial effects with possible social class effects.

The two teachers studied at the predominantly middle class school were both white females with several years teaching experience. The three teachers studied at the predominantly lower class school included two white females with several years teaching experience and one black male in his second year of teaching. The teachers were included because they happened to be teaching in the fifth grade at the two schools assigned to us by the school district in response to our request for a middle class and a lower class school with predominantly white student populations. Thus the teachers were not selected on the basis of anything known about them.

The original plans called for three teachers to be studied in each school, and arrangements were made to do so when the study was begun in the fall of 1971. However, the third teacher at the predominantly middle class school (a black male in his first year of teaching experience) had to be dropped from the study and could not be replaced. We were only able to get a few observations in his classroom because he was absent more than he was present during the fall semester due to illness and other problems, and he eventually resigned his teaching position. However, the resignation did not occur until late in the semester, so that it was too late to replace him and begin studying another teacher. Data collected in a different teacher's classroom would no longer be comparable with that collected in the other classrooms, since the nature of classroom interaction changes over time as the school year progresses. Thus because of this irreplaceable loss we were left with only five teachers in the study rather than six as had been originally planned.

Observation System

Classroom observations were made with a version of the Brophy-Good Dyadic Interaction Observation System (Brophy and Good, 1970b). The term "a version" was used in the previous sentence because the system contains a large number of variables, only some of which would ordinarily be included in a given study, depending upon the aims of the study and the grade level at which the research is conducted. For example, aspects of the system designed to measure teacher-child interaction during small group reading instruction were not used in the present research, since small group reading instruction did not occur in these fifth grade classrooms. At the same time, categories for variables such as student-initiated comments and questions and for teacher questions eliciting student opinion were used in the present study. These interactions do occur at the fifth grade level, whereas they are not used in studying first grade classrooms because they rarely if ever occur at that level.

The system is designed to code every dyadic interaction that occurs between the teacher and each single individual student. These include public response opportunities, in which the teacher asks a question and the student makes a response in front of the entire class, as well as private interactions concerning the student's seatwork or homework (work-related interactions) or concerning matters of classroom management or personal concerns (procedural interactions). Interactions involving praise for good behavior or warnings or criticism for misbehavior (behavioral interactions) are also coded.

Public response opportunities are coded as direct (teacher names the student before asking the question or calls on a non-volunteer), open (teacher calls on a volunteer with his hand raised), or call-out (a student calls out the answer before the teacher has a chance to select a respondent). This coding allows assessment of the degree to which the teacher or the student is primarily responsible for the number of response opportunities that a given student receives.

Teacher questions are coded as process, product, or choice questions when they deal with matters pertaining to the academic curriculum. Process questions require the student to explain a complex phenomenon or describe the steps (process) involved in arriving at an answer to a complex question. Product questions require only short answers, primarily recalling factual material from memory. Choice questions require only that the student choose among alternatives that the teacher provides in the question (yes-no questions, either-or questions, and questions that allow the student to point to or select from a set of alternatives). Thus in general process questions are the most difficult and choice questions the least difficult, although there are exceptions.

There are also two other categories of teacher questions: opinion questions and self-reference questions. Opinion questions require the student to state his opinion on some matter, which may or may not be re-

lated to the curriculum. In any case, they do not allow for determination of answers as being either correct or incorrect; they involve matters of values or opinions which do not have any single or obvious correct answer. Self-reference questions refer to matters of personal concern to the student, such as his interests, preferences, or likes and dislikes. These have no relationship to the curriculum, although they occasionally are used to introduce a topic ("Do you like oranges?... Well, today we are going to learn about oranges.").

All five types of response opportunities are coded as to whether they were direct questions, open questions, or call outs. In addition, those response opportunities which involve matters of direct reference to the curriculum (process questions, product questions, and choice questions) are also coded for the quality of the students' responses and the kinds of feedback that teachers made to these responses.

Students' responses were coded as being correct, incomplete or part correct, incorrect, "don't know" (the student says "I don't know" out loud), or no response (the student says nothing). The teacher's reaction was used in judging whether responses should be coded as correct, part correct, or incorrect. If a teacher accepted a response and treated it as correct, it was coded as correct, and this same criterion was used in coding responses as part correct or as incorrect.

Teachers' feedback reactions following students' responses (or failures to respond) were also coded (and were later tabulated separately depending upon the kind of student response which they followed). Teacher reactions coded included praise, criticism, failure to give any feedback at all, giving process feedback (giving an extended explanation), giving the answer (product feedback), calling on another student to give the answer, repeating the question, rephrasing the question or giving a clue, asking a new question, or asking the student to expand his answer. The first two of these categories (praise and criticism) involve teachers' evaluative reactions to student responses. The next four categories involve the method used by the teacher to give the student the correct answer when he has been wrong, as well as the quality of the feedback he receives (no feedback vs. product feedback vs. process feedback). The last four categories involve staying with the student, persisting in trying to get a response or in trying to get him to improve on the response given initially.

In addition to the above, teacher praise or criticism was also coded when it occurred in connection with a student's classroom behavior (during behavior contacts) or in connection with his seatwork or homework (when it occurred during work-related, private interactions).

Thus the system includes a variety of measures tapping both the quantity and quality of teacher-student interaction. Separate data are kept for each student simply by assigning all students in the class a different number, and using these numbers when recording teacher-student interaction. The numbers are then used later when collating the data, to compile separately information on the teacher's dyadic

interaction patterns with each different student in the classroom. Complete details about the system and its use are given in an extensive manual prepared for use by other researchers (Brophy and Good, 1970b).

Coder Training

Some of the coders used in this research were already experienced in using the system, while others had to be trained from scratch. Coders first read the manual and discussed it with the author, then wrote out their own examples of each of the behavioral categories, then practiced coding specially prepared videotapes used for coder training until they reached satisfactory performance criteria, and then practiced coding in the classrooms in which they were to work, continuing until they reached satisfactory performance criteria. During the latter two phases of coder training, the coders worked in pairs (with experienced coders being paired with new coders) and continued practice coding until 80% inter-coder agreement was reached.

Percentages were derived by taking the number of times that both coders coded an interaction and agreed and dividing this number by the sum of itself plus the number of times that they coded and disagreed plus the number of times that Coder A coded and Coder B did not plus the number of times that Coder B coded and Coder A did not. Thus the 80% figure, as defined, is a rather strict criterion of inter-coder agreement. Coders took from one to three weeks to reach this criterion. Once they reached it they ceased working in pairs and began collecting data working individually in their assigned classrooms. For futher details about coder training and assessment of inter-coder agreement, see Brophy and Good (1970b).

Observation Period

It had originally been intended that classrooms would be observed for an entire morning or an entire afternoon, as had been done previously in research at the first grade level. However, it was discovered that at the fifth grade level, not only in the two schools studied, but in the entire school system involved, a degree of departmentalization through team teaching was standard procedure. This meant that at certain periods during the day each teacher received students from other classrooms for instruction in a given subject, while some of her students left for instruction from another teacher. For example, the two teachers working in the predominantly middle class school had arranged a trade-off for the subjects of language arts and mathematics. In each of these subjects each student received an instructional period in which the teacher taught a structured lesson involving public response opportunities, discussion, introduction and explanation concerning new content, etc. He also spent another period in what was called a "center," in which he worked individually on a seatwork or homework assignment or else worked on activities of his own choosing (among those available).

During these "center" periods there was no structured teaching and relatively little student-teacher interaction. One of the two teachers taught both structured lessons in language arts while the other ran the "center" during language arts periods, and then the two teachers reversed roles during the two math periods.

A similar arrangement was used at the predominantly upper lower class school, although it was more complicated because more teachers were involved and because all classes involved structured teaching (there were no "centers"). One experienced white female teacher taught a high ability language arts group and a low ability math group. The black male teacher taught a high ability math group and a low ability language arts group. The third teacher (an experienced white female) taught the middle ability groups in both subjects.

These departmental arrangements forced us to shift from our original plan to use the half day as the unit of analysis, since the teachers dealt with different groups of students during the morning or the afternoon. Thus, instead of observing for the entire morning or for the entire afternoon, we observed the teachers during class periods in which they were dealing with intact groups of students whom they saw every day (regardless of whether the students happened to be in their homeroom or not). Thus for a given teacher we might have observed her regular language arts group in the morning and her regular math group in the afternoon. The result was a data collection plan more like what would be used in a high school in which teachers dealt with a given class for 50-minute periods every day. The meaningful unit is the class that meets every day at the same time, since the teacher interacting with that particular group of students during that particular time slot represents an intact group. This is true even though the teacher teaches other classes and other students during the day, and the students attend other classes and see other teachers, or perhaps even take different classes from the same teacher, at other times during the day.

In the predominantly middle class school, we observed each of the two teachers 15 times when teaching each of their two structured classes (one teaching two language arts classes and the other teaching two math classes), and also observed each of the same two teachers 15 times when they were conducting each of their two centers (one during the language arts periods and the other during math periods). Thus we had data on a total of eight intact groups or "classes" from this school, even though only two teachers were involved. Also, the same students who formed an intact group or class during one of the structured sessions also formed an identical intact group or class during one of the centers (for example, when Teacher A taught language arts to one group, Teacher B had the other group in the language arts center; when the time came to switch classes, they simply exchanged, so that the group that had been taught language arts by Teacher A remained as an intact class in the center under Teacher B and the group that had been in the center under Teacher B remained as an intact group in the language arts class under Teacher A).

This arrangement had two disadvantages compared to our original plan of observing an entire morning or afternoon. First, it involved much additional record keeping, since separate data had to be kept for eight classes instead of only two classes. Second, it reduced the volume of data available for a given class, since our observation time per class was cut from two hours (an entire morning or afternoon) to one hour (the usual length of a language arts or math class).

However, these disadvantages due to departmentalization were compensated by some serendipitous advantages that they offered. First, since we were seeing the same teachers under four separate conditions, we would in effect be able to replicate the study on these two teachers four separate times and assess the degree of stability in their tendencies to communicate differential expectations to different students. We could also investigate whether their behavior was affected by the makeup of the group (high or low ability). Secondly, since students were observed both in different situations with the same teacher and in situations with two different teachers, we could assess the degree of stability in the kinds of interaction patterns that the students have with different teachers. For example, is a given student generally either assertive and outgoing or passive and withdrawn in all of his classes, or is he assertive in some classes and withdrawn in others? The departmentalization of the fifth grade classes allowed us to investigate these additional questions that were not planned as part of the research originally.

A similar but somewhat less ideal situation existed for the data collection in the predominantly lower class school. Here two of the teachers were observed in two different classes each (one high ability and one low ability), while the third teacher (one of the white females) could be observed in only one of her classes (middle ability) due to a combination of resource limitations and schedule problems. Thus these three teachers were observed for a total of five classes. As was the case in the other school, although to a lesser degree, the departmentalization allowed us to look at both teacher and student stability vs. variability across different situations.

In summary, the original plan of observing in a single classroom for an entire morning or afternoon had to be abandoned because of the departmentalization used in the fifth grade. This meant extra record keeping and a reduction in the average length of observation from two hours to one hour. In compensation for these disadvantages, however, the departmentalization arrangement allowed us to observe the same teacher working with different groups of students and the same students working with different teachers or with the same teacher in different situations. Thus we were able to investigate a set of questions not anticipated in planning the original study, concerning the degree of stability vs. variability in both teachers and students in regard to the quantity and quality of teacher-student interactions that occurred in their classrooms.

Observations began in November and continued until shortly before the holiday break. They then resumed and continued through the end of March. Thus most of the first semester and a part of the second semester were included.

Measuring Teachers' Expectations for Student Performance

As had been done in several previous studies, the teachers were simply asked to rank their students in order of expected achievement. The instruction sheet for these rankings is shown in Appendix A. Instructions were deliberately kept vague and general so that the teachers' own criteria would be used in making the rankings, rather than criteria imposed on them by a more detailed set of instructions. This helped to insure that the rankings would reflect the teachers' actual beliefs about the performance level that the students would attain at the end of the year.

Teachers dealing with more than one class were asked to provide a separate set of rankings for each class. It was made clear to each teacher through discussions that the rankings were to be as specific to each class as possible, so that a student who appeared in each of two different classes (language arts and math, for example) did not necessarily have to have the same rank and could conceivably have quite different ranks. Also, after discussion with the two teachers at the predominantly middle class school, it was decided that rankings for the groups in the centers would be meaningless since no formal instruction occurred during that time and the teachers would have no rational basis upon which to make such rankings. Therefore the teachers at this school were asked to rank only the two classes in which they did structured teaching.

Teachers' rankings of expected achievement of their students were not shown to coders until after all observations were completed and all data collected. In addition, teachers were asked to avoid discussing their rankings with classroom coders. This was done to preclude any possibility of coder bias during classroom observations.

Data Preparation

Data for each class were tabulated according to the standard procedures used with the Dyadic System (Brophy and Good, 1970b). In each class, teacher-student interaction data are tabulated separately for each separate student. The system yields two basic types of scores: frequency scores and percentage measures. The frequency scores reflect the number of each of the various major categories of teacher-student interactions that a given student had with his teacher. To take into account absences, totals in each category are divided by the number of observations for which the student was present, so that the frequency scores reflect the average number of contacts per observation. Thus a given student's frequency score on the measure of student-initiated work contacts was computed by totaling all of the contacts he initiated with the teacher and dividing this total by the number of observations during which the student was present. Similar procedures were followed in computing frequency scores for such variables as response opportunities, teacher-initiated work contacts, teacher- and student-initiated procedure contacts, student-initiated questions, and behavior contacts.

The frequency scores described above reflect the quantitative aspects of teacher-student interaction. The percentage measures reflect the qualitative aspects. Use of percentage measures makes it possible to compare different students and groups of students even though they may differ in the frequency of different types of contacts that they have with their teachers. For example, one percentage measure is the percentage of correct responses (number of correct responses over total number of response opportunities). A student who answered 80 out of 100 questions correctly will get the same score (.80) as the student who answers 40 out of 50 questions correctly. Thus even though the first student had twice as many response opportunities as the second, conversion of the data to a percent measure allows a direct comparison on the variable of percentage of correct responses.

Similarly, many of the percentage measures allow comparison of the ways that teachers treat different students in equivalent situations. For example, although high achievers generally have more correct responses and fewer failures than low achievers, and therefore get more praise and less criticism than low achievers, meaningful comparisons can be made between the kinds of treatment that these two groups receive from the teacher when measures such as the percentage of correct answers which are praised and the percentage of failures which are criticized are used in place of frequency scores for praise and criticism.

Percentage measures always are computed by placing a subset in the numerator and then dividing by a total set which includes the numerator plus other subsets that make up the total set. As a result, percentage measures vary from 0.00 to 1.00, but never exceed 1.00. Such percentage measures are used in preference to ratio measures (in which one subset is divided by another subset), because they are confined to the range from 0.00 to 1.00, are more easily comparable with one another, and are less variable and generally less affected by variability in the denominator term.

For example, in dealing with misbehavior a teacher may merely warn a child about his behavior or she may react more intensively and angrily, to the point of personal criticism. The percentage measure used to reflect a given teacher's responses to a given student's misbehavior is warnings divided by warnings plus criticisms, as opposed to warnings divided by criticisms. Inclusion of warnings in both the numerator and the denominator in the percentage measure insures that the upper range cannot exceed 1.00, and thus avoids the extreme variability that would result if a warnings divided by criticisms ratio measure were used.

Similar procedures (dividing a subset by the total set which includes the subset in the numerator) are used in computing all of the percentage measures. Similarly, all of the frequency scores are computed by dividing the totals by the number of times that the child was present for observation. Thus all scores used are corrected for absences and for differences in the frequencies of various types of interactions that different students have with their teachers.

The present study included many students who were seen in more than one class. Data for these students were separately tabulated, and frequency scores and percentage measures were separately computed, for each different class in which they were observed. Thus, as mentioned above, the essential unit for analysis in this research was a given teacher with a given group of students who met as an intact class at the same time each day. Thus data were compiled for five classes involving structured teaching at the predominantly lower class school, and for four classes involving structure teaching and four "center" classes at the predominantly middle class school.

To summarize, frequency scores and percentage measures were calculated separately for the teachers' dyadic interactions with each individual student in each of these 13 classes. In addition, the data were also separately tabulated according to whether they came from the first half or the second half of the set of observations for each class. Thus for a given frequency score or percentage measure, a given child in a given class had two scores, one for the first half and one for the second half of the total number of times his class was observed.

This division of the data into first half and second half sets was done to investigate the polarization hypothesis (i.e., to see whether highs and lows were more different from each other in the second half data than in the first half data).

Data Analyses

Tests for replication data and for data to evaluate the polarization hypothesis were accomplished with analyses of variance, while questions regarding the stability of teacher and student behavior across different situations were addressed with correlational analyses. For the analyses of variance, the expectation rankings in each classroom were divided into high, middle, and low thirds, with any extra children going into the middle third.

Original plans called for teacher (6) by sex of student (2) by expectation level (3) analyses of variance over repeated measures, with teacher, sex of student, and expectancy level as between subjects independent variables and trials (first half vs. second half data) as a within subject independent variable. The dependent variables were all of the frequency scores and percentage measures from the Dyadic System. Replication data would be analyzed through analyses of the main effects and interactions involving teacher, sex of student, and (especially) expectancy level. Assuming that evidence of expectancy effects were found, the polarization hypothesis would be evaluated by analyzing the interactions between expectancy level and trials (first half vs. second half), to see if there were evidence of polarization over time.

Although the same basic logic was retained in approaching the data, these original analysis plans were changed in response to certain limitations in the present data and to experience gained in the meantime in other studies involving data from the Dyadic System.

Limitations in the present data included very low frequencies in certain classrooms on some measures of interest, as well as the fact that the teachers' expectancy groups often were very unevenly distributed between the two sexes. In the most extreme case, one teacher named only boys to the high group in her classroom.

These data limitations forced us to abandon both teacher and sex of student as independent variables. Empty cells altogether prevented the possibility of analyses of variance using sex of student, and analyses using teacher as an independent variable were unsatisfactory because empty cells prevent analyses of too many dependent variables of interest. Therefore, the school was used as an independent variable, with data from the five classes at the predominantly lower class school and data from the four structured classes in the predominantly middle class school being combined and treated as single sets.

Data from the four "center" classes in the predominantly middle class school were not included in the analyses of variance, because they differed qualitatively from the structured classes. They contained no public response opportunities because no structured teaching was done at these times. They did contain student- and teacher-initiated work and procedural contacts, as well as behavioral contacts, so that they were included in the correlational analyses.

Another modification of the original data analysis plan was decided upon on the basis of experience from several studies using high, middle, and low expectancy groups in analyses of variance. Data from several studies (Brophy and Good, 1973) show that the middle group almost invariably lies between the high and the low group on all measures. It is sometimes closer to the high group and sometimes closer to the low group, but almost always in between. These results suggested dropping the middle group out of the analyses of variance, retaining only the high and low groups.

This adjustment makes for greater simplicity in testing for expectation effects, since a significant main effect for expectancy when only two expectancy groups are used constitutes a significant difference between the groups. In contrast, a significant main effect when three groups are used shows only that the expectancy grouping had an effect. Statements about the significance (or lack of significance) of differences between any two of the three groups cannot be made without further statistical tests. Thus although the question of the relative position of the middle group compared to the high and low group on a given variable is of some minor interest, tests for replication of earlier results and for the possibility of polarization effects could be done much more simply if only the high and the low groups were included in the analyses.

Thus, combining the above considerations, the data were first approached with a school (2) by expectancy level (2) analysis of variance over repeated measures (2), the latter consisting of the

first half vs. the second half data. Data from these analyses will be discussed in the first part of the results section below.

Even with this simplified analysis of variance model, however, empty cells prevented analyses of several variables of interest. Because of this, a second set of analyses were run, using adjusted (standardized) frequency and percentage scores. Scores were standardized within each class by setting the class mean equal to zero and establishing a standard deviation of 1.00. These transformations allowed each student in a given class to retain his position relative to his classmates on each variable, but by setting each class mean equal to zero and standard deviation equal to 1.00, they allowed for combination of all of the data into a single set.

The standardization procedure eliminates teacher and school differences (class means on all variables are set equal to zero), so that this information is lost in the process. However, by allowing the combination of data from all students in all classes into one data set, standardization allowed us to reintroduce the sex of student variable and to get expectation data on the dependent variables that had had empty cells and therefore could not be analyzed in the previous set of analyses. These standardized score distributions of dependent variables were analyzed with sex of student (2) by expectancy group (2) analyses of variance over repeated measures (2), with the latter being the first half vs. the second half data. Results from these standardized score analyses will be discussed in the second part of the results section.

The third part of the results section will present correlational data reflecting the degree of stability shown by teachers and students in their interaction patterns in different classes and different kinds of class situations (structured teaching vs. "centers").

R E S U L T S

Data from the school (2) by expectancy group (2) analyses of variance over repeated measures (2) of 41 dependent interaction variables are presented in Table 1. The data are from the raw (non-standardized) percentage scores and frequency measures, and from only the nine classes involving structured teaching (not the "center" classes).

The data are divided into seven clusters. Measures within each cluster are all related to the same major aspect of teacher-student interaction. The same clusters and variable numbers will be used throughout this section in presenting the data. Where no data appear for certain variables in Table 1, one or more empty cells prevented the possibility of an analysis of variance with this model. This occurred with eight of the 49 interaction variables used.

Table 1. School and Expectancy Group Means and Probability Values from an Analysis of Variance in Raw (Non-Standardized) Interaction Measures[1]

	Group means				Probability Values[2]						
					Between subjects			Within subjects			
	School		Expectancy		School	Expectancy	School X Expectancy	Trials	School X Trials	Expectancy X Trials	School X Expectancy X Trials
	Low	Middle	High	Low							
A. Student Performance Indicators											
A1. Correct Answers/ Total Answers	.65	.80	.77	.68	<.0001	.002	.21	.66	.32	.53	.83
A2. Wrong Answers/ Wrong + Don't Know + No Response	.63	.75	.71	.67	.10	.68	.65	.40	.60	.06	.24
A3. No Response/ Don't Know + No Response	--	--	--	--	--	--	--	--	--	--	--
B. Frequencies of Each Type of Teacher-Student Interaction											
B1. Student-Initiated Questions	.04	.07	.06	.05	.11	.56	.30	.68	.23	.10	.08
B2. Reading Turns	.06	.03	.04	.04	.004	.75	.04	.19	.0002	.03	.61
B3. Direct Questions	.61	.39	.50	.50	.0005	.97	.41	<.0001	.47	.34	.75

B4. Open Questions	.51	.32	.48	.35	.002	.03	.15	.82	.19	.21	.59
B5. Call Outs	.01	.02	.01	.02	.30	.92	.60	.82	.16	.91	.16
B6. Total Academic Response Opportunities (Direct questions + open questions + call outs)	1.13	.72	.99	.86	.0001	.15	.13	.0004	.31	.90	.76
B7. Self-Reference Questions	.03	.005	.02	.02	.0003	.55	.14	.001	.01	.15	.08
B8. Opinion Questions	.03	.02	.03	.02	.47	.70	.09	.008	.0006	.12	.57
B9. Student-Initiated Work Contacts	.39	.22	.32	.28	.004	.47	.99	.76	.0001	.38	.31
B10. Student-Initiated Procedure Contacts	.22	.12	.18	.17	.02	.82	.76	.14	.05	.98	.91
B11. Total Student-Initiated Private Contacts (work + procedure)	.61	.34	.50	.45	.001	.52	.87	.48	.0001	.57	.46
B12. Teacher-Initiated Work Contacts	.15	.04	.09	.10	<.0001	.47	.64	.72	.82	.20	.31
B13. Teacher-Initiated Procedure Contacts	.21	.20	.21	.19	.67	.58	.02	.007	.0001	.27	.05
B14. Total Teacher-Initiated Private Contacts (work + procedure)	.36	.24	.30	.30	.0003	.93	.07	.02	.0006	.84	.03
B15. Warnings and Criticisms for Misbehavior	.08	.05	.05	.07	.09	.36	.06	.18	.05	.18	.11
C. Teacher vs. Student Initiation of Contacts											
C1. Total Teacher-Initiated Academic Contacts (reading turns + direct questions + teacher-initiated work contacts)	.83	.45	.64	.64	<.0001	.92	.36	<.0001	.71	.37	.97
C2. Total Student-Initiated Academic Contacts (student-initiated questions + open questions + call outs + student-initiated work contacts)	.95	.62	.87	.69	.002	.07	.29	.80	.009	.43	.56

C3. Direct Questions/ Direct + Open + Call Outs										
.53	.61	.55	.59	.05	.28	.95	.003	.10	.07	.78
C4. Open Questions/ Direct + Open + Call Outs										
.46	.38	.44	.40	.04	.28	.85	.003	.05	.05	.73
C5. Call Outs/ Direct + Open + Call Outs										
.008	.01	.01	.01	.29	.96	.23	.48	.13	.75	.67
D. Type of Teacher Questions										
D1. Process Questions/ Process + Product + Choice										
.06	.07	.07	.06	.69	.46	.71	.24	.39	.33	.63
D2. Choice Questions/ Process + Product + Choice										
.03	.08	.07	.04	.003	.04	.01	.18	.62	.67	.09
E. Teacher Praise and Criticism										
E1. Total Praise for good work or correct responses										
.02	.004	.006	.02	.002	.02	.17	.99	.03	.96	.26
E2. Praise following correct answers/ Total correct answers										
.01	.05	.03	.03	.008	.83	.74	.22	.16	.21	.17
E3. Praise following part-correct answers/ Total part-correct answers										
.000	.000	.000	.000	1.00	1.00	1.00	1.00	1.00	1.00	1.00
E4. Praise in Student-Initiated Work Contacts/ Praise + Criticism in Student-Initiated Work Contacts										
--	--	--	--	--	--	--	--	--	--	--
E5. Praise in Teacher-Initiated Work Contacts/ Praise + Criticism in Teacher-Initiated Work Contacts										
--	--	--	--	--	--	--	--	--	--	--
E6. Praise for Good Behavior										
.0004	.0000	.0000	.0004	.34	.34	.34	.34	.34	.34	.34
E7. Total Criticism for poor work or poor responses										
.02	.002	.01	.008	.004	.74	.74	.13	.008	.76	.76
E8. Criticism following incorrect answers/ Total incorrect answers										
.03	.0000	.02	.01	.28	.88	.88	.66	.66	.54	.54
E9. Criticism following "don't know" or no response/ Total "don't know" or no response										
--	--	--	--	--	--	--	--	--	--	--
E10. Warnings about misbehavior										
.04	.04	.03	.05	.59	.32	.07	.19	.04	.18	.10

E11.Criticism for Misbehavior										
=	.03	.01	.02	.02	.008	.71	.05	.22	.64	.37	.44
F. Teacher Persistence in Eliciting Responses										
F1. New Question following correct answers/ Total correct answers										
	.07	.03	.04	.06	.002	.34	.14	.45	.58	.52	.76
F2. Expansion questions following correct answers/ Total correct answers										
	.001	.001	.002	.0006	.92	.22	.39	.38	.24	.06	.81
F3. % Stays with student after part-correct answers										
	.31	.50	.40	.41	.28	.98	.98	.86	.86	.005	.01
F4. % Stays with student after incorrect answers										
	.12	.07	.06	.13	.47	.28	.99	.56	.84	.68	.44
F5. % Stays with student after "don't know" or no response										
	--	--	--	--	--	--	--	--	--	--	--
F6. Repeats question + Asks for expansion after part-correct answers/ Repeats question + Asks for expansion + Gives clue or help after part-correct answers										
	--	--	--	--	--	--	--	--	--	--	--
F7. Repeats question after incorrect answer/ Repeats question + Gives clue or help after incorrect answer										
	--	--	--	--	--	--	--	--	--	--	--
F8. Repeats question after "don't know" or no response/ Repeats question + Gives clue or help after "don't know" or no response										
	--	--	--	--	--	--	--	--	--	--	--
G. Level of Feedback Given to Students										
G1. Process feedback following failure to answer correctly/ Process + Product feedback after failure to answer correctly										
	.05	.008	.02	.04	.22	.66	.95	.78	.44	.83	.47
G2. Process feedback in student initiated work contacts/ Total Student-Initiated Work Contacts										
	.17	.02	.10	.09	.0005	.81	.19	.73	.94	.27	.24
G3. Process feedback in teacher-initiated work contacts/ Total teacher-initiated work contacts										
	.21	.0000	.12	.10	.21	.90	.90	.97	.97	.80	.80

G4. Gives answer following failure to answer correctly/ Gives answer + Asks other following failure to answer correctly	.09	.40	.30	.20	.0005	.21	.19	.48	.99	.64	.79
G5. Teacher failure to give feedback/ Total academic response opportunities	.08	.02	.05	.04	<.0001	.19	.54	.08	.06	.44	.55

[1]Where no data appear, analyses could not be performed because of lack of data in some cells.

[2]Probability values are from school (2) X expectancy group (2) analyses of variance over repeated measures (2), with the trials being the first and second sets of observations.

[3]No teacher praised a student after a part-correct answer. Consequently the variance was zero for this variable.

The data of primary interest to this research are the main effects and interactions involving the expectancy groups, particularly the trials by expectancy group interactions. However, both because of their intrinsic interest value and because they help set the stage for discussion of the expectancy data, the main effects for schools and trials will be discussed first.

There were several significant ($p < .05$) main effects for schools. The students in the predominantly middle class school had a higher percentage of correct answers and initiated more procedural interactions with their teachers than did the students in the predominantly lower class school. Also, the teachers in the middle class school asked a greater percentage of choice questions, were more likely to praise a correct answer, and were more likely to give an answer rather than call on another student when the originally called upon student failed to answer the question.

As far as they go, these data suggest that the middle class students were brighter and more likely to initiate interactions with the teachers (as would be expected). Also, the teacher data from the middle class school suggest a generally positive picture, with the teachers being more likely to praise correct answers and more likely to give an answer themselves rather than call on other students (thus reducing the likelihood of the development of an unhealthy competitive class atmosphere). The finding regarding choice questions was unexpected, however, since these usually easier questions would be expected to be more frequent in the lower class school. Teachers in the middle class school also had a higher mean on the measure of asking the more difficult process questions, however, but the difference was not statistically significant.

Thus, in summary, the measures on which the middle class school was significantly higher than the lower class school suggest a more optimal classroom environment in the middle class school.

This interpretation must be tempered somewhat, however, upon inspection of the measures on which the lower class school was significantly higher than the middle class school. These included most measures of response opportunities, as well as the measures of both teacher-initiated and student-initiated work contacts. Thus there were more class discussions and question-and-answer activities involving public response opportunities at the lower class school. There were also more work-related individual contacts between students and teachers, and this difference was as much due to the students as to the teachers. Thus this combination of data suggests that more work-related activities were going on in the lower class than in the middle class school.

However, it is possible that the differences on these measures occurred because the students in the middle class schools were given more individual work assignments and were able to handle them independently for longer periods of time so that they needed less frequent

contact with the teachers. If this were the case, it could be argued that the teacher behavior in both schools was appropriate to the needs and abilities of the students.

The remaining measures upon which the lower class school was significantly higher than the middle class school suggest a mixture of desirable and undesirable patterns. Teachers in the lower class school were more likely to ask open than direct questions. Some might see this as a favorable index, suggesting less teacher-domination and greater opportunity for the student to determine his response opportunities. However, in combination with the tendency of these same teachers to call on another student when the first student could not answer a question, the data suggest that the teachers in the lower class school may have been fostering a highly competitive, perhaps destructively competitive, group atmosphere.

The findings that these teachers were more likely to ask a new question following a correct response and also more likely to fail to give feedback following a student's response also tie in with this interpretation. In sum, these data suggest that the teachers in the lower class school were "right answer" oriented, at least during discussions and question and answer sessions. Their data suggest a pattern of going from one student to another until they got the answer they were seeking, and also a pattern of staying with a student who tended to give the right or desired answer. Again, this kind of teacher behavior makes for unhealthy competitiveness in the classroom.

Similarly mixed findings show up in the remaining measures on which the lower class school had higher means. Teachers at the lower class school gave more total praise to their students, but they also criticized them more often, both for work-related failures and for misbehaviors. They also gave their students more process feedback in student-initiated work contacts, suggesting a concerted attempt to work with the students to help them understand the material, but at the same time they more frequently failed to give feedback to the students following their responses in public response opportunities.

In sum, the school difference data suggest that teachers in the two schools were behaving in ways generally appropriate to the student populations they served, but that for the most part the learning environment in the middle class school was superior to that in the lower class school. Several measures taken in combination suggest that the teachers in the lower class school probably were creating a highly competitive and perhaps destructively competitive group atmosphere in their classrooms, in comparison with the classrooms in the middle class school.

In addition to the school effects, there were also several significant trials effects reflecting changes in the kinds of activities that occur in the classroom as the school year progresses. The

frequencies of self-reference and opinion questions went down from the first to the second set of data, while the frequencies of direct questions, direct questions over direct plus open questions, total public response opportunities, and teacher-initiated work and procedure contacts all went up. These data show that as the school year progresses the classroom interaction becomes more and more focused on matters of direct relevance to the curriculum (fewer opinion and self-reference questions), and that the teachers become more proactive or directive in managing instruction (more teacher-initiated individual contacts, more direct questions).

Inspection of the several schools by trials interactions which appeared showed that most of the change that occurred between the first and the second set of data occurred in the middle class school. Teachers in this school began to ask notably more direct questions and to initiate more private interactions with students during the second half. Interestingly, their students also began to initiate work-related interactions more frequently with them. These interaction data further support the suggestion that the learning environment at the middle class school was superior to that at the lower class school. The only negative note was an interaction for the measure of behavioral warnings and behavioral criticisms. These increased notably in the middle class school while remaining constant in the lower class school. Still, however, behavioral warnings and criticisms were not as frequent in the middle class school even in the second half of the data as they were in the lower class school.

We turn now to the data on the high and low expectancy groups, to see whether findings from the first grade level (Brophy and Good, 1970a) were replicated in these data and to assess the polarization hypothesis as it applies to the expectancy groups by trials interaction data.

In general, the findings for expectancy groups were weak in intensity and mixed in direction. Only four of the 41 dependent measures yielded significant ($p < .05$) main effects for expectancy groups, even though only the high and low groups were included in the analyses. The highs had more correct answers per response opportunity than the lows, and they also answered more open questions than the lows (indicating that they raised their hands more frequently and were called on more frequently in these situations by the teachers). These findings confirm previous results, although they deal with student-initiated behavior and are not to be taken as evidence of communication of expectations by the teachers.

The other two significant main effects show a reversal of previous findings: the lows received more total praise than the highs, and the highs answered a greater percentage of the easier choice questions than did the lows (the highs also answered a higher percentage of the more difficult process questions, but the differ-

ence here was not statistically significant). Thus in general the Brophy and Good (1970a) findings from the first grade level were not replicated in the present study. Only four of a possible 41 measures showed a significant expectancy group effect, and only two of these differences were in the hypothesized direction.

The data could be interpreted as providing support, but very weak support, for the suggestion that as children get older expectancy effects will show up in quantitative but not qualitative measures. Most of the quantitative measures (Cluster B) show that the means for the highs equal or exceed those for the lows, but the expectancy group effect is significant for only one of a possible 15 quantitative measures (open questions).

The nonsignificant group differences in the other data are mixed. The greater tendency of the highs to initiate work-related contacts with the teacher almost reached statistical significance ($p = .07$), and, along with the finding regarding open questions, helps confirm the previous findings that the highs are more initiatory and active than the lows in seeking and getting contacts with the teachers. However, the nonsignificant group differences for teacher praise and criticism are in the opposite direction from the Brophy and Good (1970a) findings, in that they suggest less praise and more criticism towards highs than lows (Cluster E). The same is true for the measures of teachers' persistence in seeking responses from students (Cluster F). The findings regarding the quality of feedback given to students (Cluster D) are mixed, and none reached statistical significance.

There were also four statistically significant ($p < .05$) schools by expectancy groups interaction effects. Inspection of the group means involved in these interactions showed a mixed and uninterpretable pattern. Thus there was no clear tendency for one school or the other to be more likely to show expectancy effects in these data.

Significant interactions between expectancy groups and trials were examined to assess their implications with regard to the polarization hypothesis. This was done even though the main effects for expectancy were weak, because there is evidence to suggest that expectancy effects are increasingly likely to show up as the school year goes on (Brophy and Good, 1972). However, the findings from these interaction data with regard to the polarization hypothesis are similar to the findings from the main effects data with regard to replication: very few reach statistical significance and they are mixed in direction. Only three of a possible 41 expectancy groups by trials interactions reached the .05 level of significance.

One of these interactions involved reading turns. The frequency of reading turns for the high group decreased from the first to the second half of the data, while the frequency in the low group remained constant. If anything, this would be interpreted

as contrary to the polarization hypothesis. This certainly would be the case if the data were from the first or second grade level. However, at the fifth grade level, where reading turns are infrequent, the meaning of this finding is difficult to interpret.

One possible explanation for the decrease in reading turns from the first to the second half of the data set is that the high groups finish their readers early and spend the remainder of the school year working on individual assignments. Students usually do these assignments independently at their seats and progress at their own pace. The lows, on the other hand, may continue to meet with their reading groups until the end of the school year (or until they finish their readers).

A second significant interaction concerned the measure of open questions divided by the total of direct questions plus open questions plus call outs. The means on this measure decreased from the first half to the second half of the data for both the high and the low groups, but the decrease was much more pronounced in the low group. As far as it goes, this finding is consistent with the polarization hypothesis. It suggests that the lows raised their hands less often to seek a response opportunity as the school year went on, so that they were called on less often to answer open questions. However, the teachers tended to compensate for this by calling on the lows to answer direct questions more frequently, so that the lows continued to have almost as many response opportunities as the highs during the second half of the observation period as they did in the first half. Thus, as was the case with the main effects, the positive finding here only reflects change in student behavior and cannot be interpreted as evidence of teacher communication of expectations.

The third significant interaction concerned the measure of the teachers' willingness to stay with a student following a part-correct answer to an original question (Variable F-3). The two-way interaction between expectancy groups and trials goes against the polarization hypothesis, in that the mean for lows on this measure rose and the mean for highs fell from the first to the second set of data. However, inspection of the three-way interaction involving schools as well as expectancy groups and trials showed that the effect was confined almost entirely to the middle class school. Furthermore, part-correct answers were extremely infrequent relative to the other kinds of student responses, so that only a very few instances of teacher feedback to part-correct responses were observed at either school. This calls into question the reliability of the finding, and in combination with the lack of any other clear cut findings on other measures of teacher persistence in seeking responses (Cluster F), it should not be accepted without replication.

Significant three-way interactions involving schools, expectancy groups, and trials also appeared for the measures of teacher-initiated procedure contacts (B-13) and total teacher-initiated

private contacts (B-14). Inspections of the means involved in these interactions again showed an uninterpretable pattern: the data for one school showed movement congruent with the polarization hypothesis while that for the other school showed movement in the opposite direction.

Taken together, the interaction data involving expectancy groups in the present study were mostly not statistically significant, and those that were significant were largely ambiguous or uninterpretable with regard to the polarization hypothesis. This is essentially the same set of findings that emerged in the earlier attempt to test the polarization hypothesis (Brophy and Good, 1973), and the same conclusion seems warranted: the polarization hypothesis remains essentially untested because once again the teachers involved in the study showed little or no evidence of expectancy effects, and thus provided no basis for investigating whether teacher expectancy effects will polarize the class over time.

A few findings regarding the middle expectancy group are worth noting. As expected, the middle group's mean fell in between the means for the high and low groups on most measures. However, the middle group had the highest mean on two measures and the lowest mean on nine measures. A significant group effect was involved on only one of these measures, that concerning the teacher's tendency to give the answer versus call on someone else following an initial failure (Variable G-4). The mean for the high group on this measure was .30, the mean for the low group .20, and the mean for the middle group .10. Thus teachers were especially likely to move away from the middle students and call on someone else more than they were with the other two groups. This fits in with earlier data from several studies (Brophy and Good, 1973) suggesting that the middle group of students tends to be less salient to the teacher and generally more passive in the classroom than either the high or the low groups.

The middle group means on measures where there was no significant expectancy groups effect also tend to bear out this description. The middle students had more reading turns than the other groups, and teachers were more likely to repeat the question to them than to give them help or ask a new question.

At the same time, the middle group's mean was lower for the measures of student-initiated questions, student-initiated work-related interactions, student-initiated procedural interactions, total student-initiated private interactions, total teacher-initiated private interactions, and wrong answers over wrong answers plus "don't know" plus no response. Also, teachers were least likely to stay with these students following part-right answers, as well as more likely to call on someone else than to give the answer following an initial failure. Thus once again, although with one exception, the group differences are small and not statistically significant, a pattern that has been observed for the middle group in several previous studies (Brophy and Good, 1973) emerges:

whenever the means for the middle group are not in between those for the high and low groups, they suggest that the students in the middle group are more passive and non-initiatory in their classroom behavior, and less salient to the teachers.

The implications of the data in Table 1 with regard to the three major questions posed in this study may be summarized as follows:

1. With regard to replication of the Brophy and Good (1970a) findings, the results of the present study yielded no replication on measures related to teacher expectancy effects. Even though the middle groups were excluded from the analyses, a statistically significant main effect for expectancy groups was observed on only four measures. Two of these were consistent with predictions, and two were not. None involved the key qualitative measures stressed by Brophy and Good (1970a) as indications of teacher expectation effects. Thus the Brophy and Good (1970a) findings were not replicated.

2. With regard to the hypothesis that teacher expectation effects at the higher grade levels would be mediated through quantitative rather than qualitative interaction measures, the present results provide slight positive support. One of the two significant expectancy group effects that were in the predicted direction was for the measure of open questions, a quantitative measure (the other was on the measure of correct answers over total answers, which is a student measure rather than a teacher measure and therefore is not relevant to the qualitative versus quantitative teacher behavior question). Also, most of the remaining quantitative measures were in the predicted direction, although they did not reach statistical significance. At the same time, none of the previous findings on qualitative measures were replicated. As far as they go, the present data do support the idea that as children get older teacher expectation effects will be mediated more through quantitative than qualitative measures. However, this conclusion must be taken within the larger context of the generally weak and negative results regarding expectation in the present study.

3. With regard to the polarization hypothesis, the results of the present study parallel those reported in an earlier attempt to test this hypothesis (Brophy and Good, 1973). That is, the hypothesis could not be adequately tested because widespread teacher expectation effects, which must be present before the polarization hypothesis can be tested, were not observed in the present study.

Analyses of Standardized Scores

In an effort to get data on variables of interest which do not appear in Table 1 because of one or more empty cells, and to take into account student sex, which could not be included in Table 1 because of an empty cell, two sets of analyses of variance in standard scores were performed. Students' scores on each of the 49 classroom observation measures were standardized within each class or center by setting the mean of the class or center equal to zero and establishing a standard deviation equal to 1.00. This procedure eliminated mean differences between classes and set the classes on a common scale so that data could be combined. The data were combined into two sets, one for the nine regular classes and one for the four center classes. For each of these two sets of data, expectancy groups (2) by sex of student (2) analyses of variance over repeated measures (2), with the latter being the first half and the second half of the data set, were performed for each of the 49 variables. Data from these analyses are given in Table 2 (for the regular classes) and Table 3 (for the center classes).

Expectancy Group Findings

The expectancy group data of Table 2 closely parallel the same data from Table 1. This is to be expected, given the relationship between the two sets of analyses. The main difference was that as a result of combining the data from the nine classes into a single set, six rather than four main effects for expectancy reached the .05 level of statistical significance. In Table 1, the high expectancy students had significantly higher scores than the low expectancy students for percentage of correct answers over total answers (A1), number of open questions (B4), and number of choice questions divided by the sum of process questions plus product questions plus choice questions (D2).

The first two of these differences were also significant in Table 2; the third difference was in the same direction but did not reach statistical significance. Two variables favoring highs over lows reached significance in Table 2 which approached but did not reach significance in Table 1: total response opportunities (B6) and total student-initiated private contacts with the teachers (C2). These two variables yielded significant main effects for schools in Table 1, indicating that the advantage of highs over lows on these two measures was primarily confined to the lower class school. Differences at the middle class school were negligible.

Two of the six significant main effects for expectancy groups in Table 2 favored the lows: lows were more often praised for good work (E1) and more often given feedback following their responses (E5). The first difference was also significant in Table 1, while the second was not. Again, the discrepancy between the two tables is due to a difference between schools. These school differences (and more particularly differences among individual teachers) will be discussed in a later section.

Table 2. Expectancy and Sex Group Means and Probability Values from Analysis of Variance in Standardized Interaction Measures (Regular Classes).[1]

	Means				Probability Values[2]							
					Between Subjects			Within Subjects				
	Expectancy		Sex		Expectancy	Sex	Expectancy X Sex	Trials	Expectancy X Trials	Sex X Trials	Expectancy X Sex X Trials
	Highs	Lows	Boys	Girls							
A. Student Performance Indicators											
A1. Correct Answers/ Total Answers	.15	-.20	-.12	.07	.004	.11	.99	.46	.82	.40	.76
A2. Wrong Answers/ Wrong + Don't Know + No Response	.10	.03	.05	.08	.66	.86	.95	.48	.19	.84	.84
A3. No Response/ Don't Know + No Response	.14	-.04	.15	-.06	.39	.31	.08	.40	.74	.98	.15
B. Frequencies of Each Type of TEacher-Student Interaction											
B1. Student-Initiated Questions	.17	.01	.21	-.03	.19	.04	.90	.77	.38	.70	.74
B2. Reading Turns	.03	-.01	.13	-.11	.69	.04	.29	.91	.06	.71	.31
B3. Direct Questions	.10	-.06	.17	-.12	.22	.02	.19	.68	.24	.76	.97
B4. Open Questions	.26	-.16	.23	-.12	.004	.01	.25	.61	.08	.66	.72
B5. Call Outs	.09	-.09	.12	-.11	.12	.05	.30	.48	.79	.54	.74
B6. Total Academic Response Opportunities (Direct Questions + Open Questions + Call Outs)	.22	-.13	.24	-.15	.01	.006	.17	.79	.88	.96	.41
B7. Self-Reference Questions	-.04	-.01	-.06	.004	.77	.46	.02	.84	.83	.76	.94
B8. Opinion Questions	-.02	.002	.02	-.04	.82	.51	.73	.78	.29	.79	.71
B9. Student-Initiated Work Contacts	.13	.03	-.08	.25	.49	.02	.28	.43	.23	.32	.52

B10. Student-Initiated Procedure Contacts										
.10	.03	.02	.11	.58	.51	.54	.84	.65	.26	.09
B11. Total Student-Initiated Private Contacts (Work + Procedure)										
.14	.03	-.04	.22	.42	.07	.77	.49	.33	.94	.26
B12. Teacher-Initiated Work Contacts										
.02	.13	.21	-.06	.40	.03	.67	.47	.35	.70	.65
B13. Teacher-Initiated Procedure Contacts										
.16	-.05	.21	-.09	.10	.02	.03	.73	.37	.05	.87
B14. Total Teacher-Initiated Private Contacts (Work + Procedure)										
.13	.03	.25	-.09	.40	.006	.02	.99	.99	.09	.90
B15. Warnings and Criticisms for Misbehavior										
-.007	.003	.12	-.13	.93	.04	.83	.43	.70	.25	.21
C. Teacher vs. Student Initiation of Contacts										
C1. Total Teacher-Initiated Academic Contacts (Reading Turns + Direct Questions + Teacher-Initiated Work Contacts)										
.08	.01	.24	-.15	.62	.002	.12	.72	.38	.84	.99
C2. Total Student-Initiated Academic Contacts (Student-Initiated Questions + Open Questions + Call Outs + Student-Initiated Work Contacts)										
.29	-.09	.15	.06	.01	.56	.24	.86	.27	.07	.71
C3. Direct Questions/ Direct + Open + Call Outs										
-.09	.08	.07	.06	.22	.35	.86	.69	.01	.98	.65
C4. Open Questions/ Direct + Open + Call Outs										
.10	-.08	.06	-.03	.21	.52	.83	.67	.01	.95	.63
C5. Call Outs/ Direct + Open + Call Outs										
.06	-.08	.11	-.12	.22	.04	.34	.66	.79	.73	.46
D. Type of Teacher Questions										
D1. Process Questions/ Process + Product + Choice										
.05	.03	.13	-.04	.85	.16	.68	.95	.22	.26	.46
D2. Choice Questions/ Process + Product + Choice										
.07	-.02	-.03	.08	.36	.27	.06	.72	.21	.74	.90

E. Teacher Praise and Criticism										
E1. Total Praise for Good Work or Correct Responses										
-.07	.15	.05	.02	.02	.76	.80	.50	.45	.46	.95
E2. Praise Following Correct Answers/ Total Correct Answers										
.04	-.01	.04	-.01	.62	.62	.04	.82	.30	.88	.86
E3 Praise Following Part-Correct Answers/ Total Part-Correct Answers										
0	0	0	0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
E4. Praise in Student-Initiated Work-Contacts/ Praise + Criticism in Student-Initiated Work Contacts										
--	--	--	--	--	--	--	--	--	--	--
E5. Praise in Teacher-Initiated Work Contacts/ Praise + Criticism in Teacher-Initiated Work Contacts										
--	--	--	--	--	--	--	--	--	--	--
E6. Praise for Good Behavior										
-.03	-.008	-.03	-.009	.43	.48	.14	.81	.36	.37	.36
E7. Total Criticism for Poor Work or Poor Responses										
-.04	-.03	-.07	-.003	.95	.35	.28	.42	.46	.61	.98
E8. Criticism Following Incorrect Answers/ Total Incorrect Answers										
-.03	-.02	.06	-.12	.94	.12	.63	.78	.13	.09	.13
E9. Criticism Following "Don't Know" or No Response/ Total "Don't Know" or No Response										
-.22	.08	.02	-.16	.06	.26	.80	.78	.56	.84	.69
E10. Warnings about Misbehavior										
.003	-.02	.14	-.15	.84	.01	.86	.44	.40	.19	.51
E11. Criticism for Misbehavior										
-.02	-.003	.06	-.09	.86	.13	.46	.25	.89	.91	.11
F. Teacher Persistence in Eliciting Responses										
F1. New Questions Following Correct Answers/ Total Correct Answers										
-.04	.05	.14	-.13	.48	.02	.41	.51	.54	.06	.96

F2.	Expansion Questions Following Correct Answers/ Total Correct Answers									
	.04	-.02	-.02	.04	.34	.34	.40	.82	.08	.34	.68
F3	% Stays with Student After Part-Correct Answers									
	-.24	.02	-.09	-.13	.26	.87	.003	.53	.45	.17	.13
F4.	% Stays with Student After Incorrect Answers									
	-.11	.03	-.02	-.07	.37	.75	.73	.62	.65	.82	.18
F5.	% Stays with Student After "Don't Know" or No Response									
	-.07	-.18	-.15	-.10	.54	.78	.16	.33	.01	.36	.95
F6.	Repeats Question After Part-Correct Answer + Asks for Expansion/ Repeats + Asks for Expansion + Gives Clue or Help After Part-Correct Answers									
	--	--	--	--	--	--	--	--	--	--	--
F7.	Repeats Question After Incorrect Answer/ Repeats + Gives Clue or Help After Incorrect Answer									
	-.05	.22	.26	-.09	.68	.60	.69	.39	.64	.17	.12
F8.	Repeats Question after "Don't Know" or No Response/ Repeats + Gives Clue or Help After "Don't Know" or No Response									
	--	--	--	--	--	--	--	--	--	--	--
G.	Level of Feedback Given to Students									
G1.	Process Feedback Following Failure to Answer Correctly/ Process + Product Feedback Following Failure to Answer Correctly									
	.04	.03	.08	-.005	.96	.57	.97	.79	.68	.96	.84
G2.	Process Feedback in Student-Initiated Work Contacts/ Total Student-Initiated Work Contacts									
	.11	.06	.009	.16	.77	.32	.003	.82	.08	.21	.72
G3.	Process Feedback in Teacher-Initiated Work Contacts/ Total Teacher-Initiated Work Contacts									
	.30	-.05	.03	.22	.11	.36	.60	.54	.70	.14	.08
G4.	Gives Answer Following Failure to Answer Correctly/ Gives Answer + Asks Other After Failure to Answer Correctly									
	.07	.01	.07	.02	.73	.78	.04	.80	.47	.39	.15

G5. Teacher Failure to Give Feedback/ Total Academic Response Opportunities	.07	-.18	-.11	-.006	.02	.36	.55	.70	.04	.76	.99

[1]Where no data appear, analyses could not be done because of lack of data in some cells.

[2]Probability values are from expectancy group (2) X sex (2) analyses of variance over repeated measures (2), with the trials being the first and second set of observations.

The two significant expectancy groups by trials interactions in Table 2 reflect the same point made earlier: as the school year went on, the teachers began calling on the low expectation students less frequently for open questions and more frequently for direct questions.

Although there are some minor differences, the data of Table 2 suggest the same conclusions regarding expectation effects in the nine regular classrooms as did the data in Table 1. Expectancy groups were not an important factor affecting teacher-student interaction patterns in these classrooms, and few of the Brophy and Good (1970a) findings were replicated. The few significant differences which did appear are consistent with the hypothesis that expectation effects at the fifth grade level are mediated largely through quantitative rather than qualitative measures of teacher-student interaction, but this conclusion must be viewed within the larger context of weak and mostly negative findings.

Even after standardization of scores to combine them into a single set, four of the 49 teacher-student interaction measures in Table 2 still had empty cells which prevented analyses of variance. Two of these occurred because praise and especially criticism were very infrequent in private work-related contacts, and the other two occurred because part-correct answers and "don't know" and no response answers were very infrequent so that certain aspects of teacher feedback reactions following such responses could not be coded often enough to allow analyses to be performed.

Data from the four center classes are presented in Table 3. Even after standardization of scores, analyses of variance could be performed on only 34 of the 49 teacher-student interaction measures because of the low frequency of some types of teacher-student interactions in these center classes. Like the data for the nine regular classes, the data for the four center classes provide little support for the hypotheses studied. Only two main effects for expectancy groups reached statistical significance. The teachers initiated more work-related interactions in the centers with the low expectation students (B12), and they also criticized the low expectation students more frequently for misbehavior (F11). Although the latter results replicate earlier findings (Brophy and Good, 1970a), the former goes against the hypothesis. Even though it is a quantitative measure of classroom interaction, it favors the lows over the highs. Teachers more frequently initiated work-related interactions with the lows, suggesting that they not only were unaffected by negative expectations but were making a concerted effort to work with these low expectation students.

Only one expectancy group by trials interaction reached significance: as the school year progressed, high expectation students increased and low expectation students decreased the frequency with which they initiated procedural interaction with their teachers.

Table 3. Expectancy and Sex Group Means and Probability Values from Analyses of Variance in Standardized Interaction Measures (Center Classes).[1]

	Means				Probability Values[2]							
	Expectancy		Sex		Between Subjects			Within Subjects				
	Highs	Lows	Boys	Girls	Expectancy	Sex	Expectancy X Sex	Trials	Expectancy X Trials	Sex X Trials	Expectancy X Sex X Trials	
A.	Student Performance Indicators											
A1.	Correct Answers/ Total Answers											
	-.02	-.03	.30	-.36	.98	.08	.35	.67	.77	.72	.87	
A2.	Wrong Answers/ Wrong + "Don't Know" + No Response											
	--	--	--	--	--	--	--	--	--	--	--	
A3.	No Response/ "Don't Know" + No Response											
	--	--	--	--	--	--	--	--	--	--	--	
B.	Frequencies of Each Type of Teacher-Student Interaction											
B1.	Student-Initiated Questions											
	.02	.05	.03	.04	.86	.96	.64	.63	.69	.16	.58	
B2.	Reading Turns[3]											
	.00	.00	.00	.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
B3.	Direct Questions											
	.06	-.05	.08	-.07	.52	.35	.54	.06	.56	.19	.51	
B4.	Open Questions											
	-.03	-.002	-.04	.007	.87	.79	.64	.52	.30	.02	.91	
B5.	Call Outs											
	-.06	.16	.12	-.02	.08	.26	.45	.63	.41	.07	.08	
B6.	Total Academic Response Opportunities (Direct + Open + Call Outs)											
	-.05	.01	-.01	-.03	.72	.92	.75	.61	.42	.07	.43	
B7,	Self-Reference Questions[3]											
	.00	.00	.00	.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	

B.8	Opinion Questions										
	.00	.00	.00	.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
B9,	Student-Initiated Work Contacts										
	.14	-.10	+.02	.07	.28	.69	.41	.44	.90	.80	.86
B10.	Student-Initiated Procedure Contacts										
	.04	.02	+.03	.03	.92	.97	.92	.60	.02	.60	.78
B11.	Total Student-Initiated Private Contacts (Work + Procedure)										
	.12	-.06	+.01	.05	.45	.87	.70	.46	.15	.99	.89
B12.	Teacher-Initiated Work Contacts										
	-.15	.27	+.13	.01	.03	.46	.33	.97	.90	.78	.78
B13.	Teacher-Initiated Procedure Contacts										
	.02	.12	+.32	-.28	.59	.01	.38	.63	.79	.96	.96
B14.	Total Teacher-Initiated Private Contacts (work + Procedure)										
	-.05	.24	+.36	-.17	.11	.005	.20	.80	.99	.92	.75
B15.	Warnings and Criticisms for Misbehavior										
	.19	.22	+.68	-.28	.85	<.0001	.95	.83	.93	.51	.66
C.	Teacher vs. Student Initiation of Contacts										
C1.	Total Teacher-Initiated Academic Contacts (Reading Turns + Direct Questions + Teacher-Initiated Work Contacts)										
	-.07	.20	+.17	-.05	.15	.23	.34	.24	.81	.35	.31
C2.	Total Student-Initiated Academic Contacts (Student-Initiated Questions + Open Questions + Call Outs + Student-Initiated Work Contacts)										
	.14	-.07	+.01	.06	.36	.84	.37	.35	.72	.57	.79
C3.	Direct Questions/ Direct + Open + Call Outs										
	.15	-.05	+.06	.04	.54	.95	.06	.12	.59	.01	.71
C4.	Open Questions/ Direct + Open + Call Outs										
	-.03	-.04	+.03	-.11	.97	.69	.08	.13	.79	.03	.62

C5.	Call Outs/ Direct + Open + Call Outs									
	-.15	.28	.14	-.01	.21	.64	.54	.84	.57	.20	.21
D.	Type of Teacher Questions									
D1.	Process Questions/ Process + Product + Choice									
	-.09	-.12	-.09	-.12	.75	.77	.21	.0004	.07	.08	.53
D2.	Choice Questions/ Process + Product + Choice									
	-.12	-.29	-.16	-.25	.43	.67	.15	.23	.22	.34	.19
E.	Teacher Praise and Criticism									
E1.	Total Praise for Good Work or Correct Responses									
	.06	.04	.005	.10	.89	.60	.18	.73	.74	.31	.52
E2.	Praise Following Correct Answers/ Total Correct Answers									
	-.02	-.09	-.04	-.07	.07	.39	.04	.007	.07	.39	.04
E3.	Praise Following Part-Correct Answers/ Total Part-Correct Answers									
	--	--	--	--	--	--	--	--	--	--	--
E4.	Praise in Student-Initiated Work Contacts/ Praise + Criticism in Student-Initiated Work Contacts									
	--	--	--	--	--	--	--	--	--	--	--
E5.	Praise in Teacher-Initiated Work Contacts/ Praise + Criticism in Teacher-Initiated Work Contacts									
	--	--	--	--	--	--	--	--	--	--	--
E6.	Praise for Good Behavior									
	.02	-.02	-.03	.03	.62	.47	.33	.96	.62	.47	.33
E7.	Total Criticism for Poor Work or Poor Responses									
	.006	.04	.04	.005	.78	.77	.30	.71	.78	.77	.30
E8.	Criticism Following Incorrect Answers/ Total Incorrect Answers									
	--	--	--	--	--	--	--	--	--	--	--
E9.	Criticism Following "Don't Know" or No Response/ Total "Don't Know" or No Response									
	--	--	--	--	--	--	--	--	--	--	--
E10.	Warnings about Misbehavior									
	.28	.06	.60	-.26	.20	< .0001	.21	.69	.28	.28	.13
E11.	Criticism for Misbehavior									
	-.08	.32	.34	-.11	.02	.008	.03	.95	.15	.94	.42

F.	Teacher Persistence in Eliciting Responses									
F1.	New Questions Following Correct Answers/ Total Correct Answers									
	.21	-.07	-.13	.27	.40	.23	.42	.67	.40	.23	.42
F2.	Expansion Questions Following Correct Answers/ Total Correct Answers[3]									
	.00	.00	.00	.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
F3.	% Stays with Student after Part-Correct Answers									
	--	--	--	--	--	--	--	--	--	--	--
F4.	% Stays with Student after Incorrect Answers									
	--	--	--	--	--	--	--	--	--	--	--
F5.	% Stays with Student after "Don't Know" or No Response									
	--	--	--	--	--	--	--	--	--	--	--
F6.	Repeats Question after Part-Correct Answers + Asks for Expansion/ Repeats + Asks for Expansion + Gives Help or Clue after Part-Correct Answers									
	--	--	--	--	--	--	--	--	--	--	--
F7.	Repeats Question after Incorrect Answer/ Repeats Question + Gives Clue or Help after Incorrect Answer									
	--	--	--	--	--	--	--	--	--	--	--
F8.	Repeats Question after "Don't Know" or No Response/ Repeats Question + Gives Clue or Help after "Don't Know"or No Response									
	--	--	--	--	--	--	--	--	--	--	--
G.	Level of Feedback Given to Students									
G1.	Process Feedback Following Failure to Answer Correctly/ Process + Product Feedback following Failure to Answer Correctly									
	--	--	--	--	--	--	--	--	--	--	--
G2.	Process Feedback in Student-Initiated Work Contacts/ Total Student-Initiated Work Contacts									
	.17	-.12	.03	.02	.27	.96	.42	.31	.75	.14	.23
G3.	Process Feedback in Teacher-Initiated Work Contacts/ Total Teacher-Initiated Work Contacts									
	-.10	.02	-.08	.00	.76	.83	.76	.58	.96	.58	.96

G4. Gives Answer Following Failure to Answer Correctly/ Gives Answer + Asks Other after Failure to Answer Correctly										
--	--	--	--	--	--	--	--	--	--	--
G5. Teacher Failure to Give Feedback/ Total Academic Response Opportunities										
.10	.03	-.09	.21	.79	.28	.73	.85	.33	.91	.33

[1]Where no data appear, analyses could not be done because of lack of data in some cells.

[2]Probability values are from expectancy groups (2) X sex (2) analyses of variance over repeated measures (2), with the trials being the first and second sets of observations.

[3]On these measures cell scores were uniformly zero, consequently the variance was also zero.

s finding fits the polarization hypothesis, but it provides little support given the variable measured and the fact that it was the only one of 34 possible interactions to reach significance.

Thus, in summary, the data from the center classes yield the same conclusions as the data from the regular classes: expectancy groups were not an important factor affecting teacher-student interaction in these classes, and the significant differences that did emerge were mixed in direction and do not support the Brophy and Good (1970a) findings or the hypothesis that the classroom would become polarized as the school year progressed.

Findings Regarding Sex of Student

A number of significant main effects for sex appear in the data of Tables 2 and 3. These findings are quite consistent with earlier results from a number of studies (summarized in Brophy and Good, 1973) showing that boys are more active in the classroom and more salient to the teachers, so that they have more frequent interactions with teachers, especially public response opportunities, and are more frequently criticized for misbehavior.

Fourteen of the 16 main effects for sex in Table 2 (regular classes) favored boys over the girls. These differences showed that the boys had more of most types of public response opportunities, that they called out answers more frequently, that the teachers initiated more contacts with them, that they were more frequently warned or criticized for misbehavior, and that they were more frequently asked a new question if they responded correctly to an initial question. In addition, all five of the significant main effects for sex in Table 3 (centers) favored boys, reflecting that the teachers initiated more private contacts with them (especially procedural contacts) and warned or criticized them more frequently for misbehavior.

All of these findings are consistent with previous research showing that boys are more active than girls in the classroom and therefore more salient to teachers. This greater activity includes sanctioned behavior (hand-raising in open question situations; student-initiated questions) as well as various forms of non-sanctioned behavior related to high activity levels and poor impulse control (calling out answers without prior recognition; misbehaving more frequently and intensively).

The two differences favoring girls in Table 2 showed that the girls initiated more work-related contacts (B9) and consequently also had a larger total number of student-initiated private contacts with the teachers (B11) in the regular classes. Again, these findings are consistent with the general literature on sex differences in elementary students' interactions with teachers. When measures of quantity of interactions with the teachers do favor girls, they almost invariably are of the present type: student-initiated private interactions

(Brophy and Good, 1973). The greater activity and saliency of boys almost invariably causes them to dominate the public response opportunities in classrooms, and measures of teacher-initiated private contacts with students almost always favor boys over girls also (the reasons for the latter findings are not yet clear; they probably include the fact that boys are more salient and therefore more likely to be objects of teacher awareness and concern, and the greater likelihood that teachers will initiate contacts with boys for control or accountability purposes). In any case, when quantitative measures of teacher-student interaction are found to favor girls, they almost always involve student-initiated private interactions rather than public response opportunities or teacher-initiated private interactions, as is evident in the data of Table 2. The girls also had higher means than boys on the measures of student-initiated private contacts in the center classes (Table 3), but here the group differences did not reach the .05 level of statistical significance.

A significant sex of student by trials effect showed up for only one of the variables in Table 2: teacher-initiated procedural contacts (B13). Inspection of the means involved in this interaction showed that the teachers were much more likely to initiate such contacts with boys than with girls early in the year. This disparity was reduced somewhat as the year went on, although the boys' mean was still higher in the second half of the data set. The facts that only this one sex by trials interaction reached significance, and that none of the three-way interactions were statistically significant, show that the sex difference data in Table 2 for the regular classes previously described held up throughout the period of observation.

Three sex by trials interactions reached statistical significance in the center classes (Table 3): open questions (E4), direct questions divided by the total of direct plus open plus call outs (C3), and open questions divided by the sum of direct plus open plus call outs (C4). All three of these significant interactions are interrelated, resulting from the same basic change over time: as the school year progressed, the boys were called on more frequently for direct questions and less frequently for open questions, while the opposite trend was seen in the girls. This may have been due to a change in student behavior, if boys began to raise their hands less frequently and/or girls to raise them more frequently in open question situations. It might also reflect a change in teacher behavior, with teachers calling on boys to answer direct questions more frequently as the year progressed. It cannot be determined from the data whether the interaction resulted from either or both of these two possible causes.

In summary, the sex difference data in both the regular classes and the center classes closely conformed to findings from several previous studies (Brophy and Good, 1973) showing that boys are more active in the classroom and more salient to the teacher than girls. Boys received more public response opportunities, more teacher-initiated private interactions, and more warnings and criticisms for misbehavior.

Girls had higher scores than boys only on the measures of student-initiated private interactions with teachers, which also is consistent with previous findings. Only a few sex by trials interactions were observed, indicating that the sex differences represented by the main effects were sustained across the portion of the school year included in the observations.

Interaction Between Expectancy Groups and Sex of Student

There were six significant sex by expectancy groups interactions in the regular class data (Table 2) and two significant sex by expectancy groups interactions in the center data (Table 3).

In the regular classes (Table 2), the interaction on the self-reference question variable (B7) occurred because the high expectation boys had more self-reference questions than the low expectation boys, while the low expectation girls had more self-reference questions than the high expectation girls. The interaction on the variable of teacher-initiated procedural contacts (B13) occurred because the high expectation boys had many more of these contacts than any of the other three groups. A sex by expectancy groups by trials three-way interaction also appeared for this variable, because the advantage of the high expectation boys was reduced somewhat as the year progressed and because both groups of girls began to receive more teacher-initiated procedural contacts as the year progressed. These two interactions showed that the main effect favoring boys on this variable was due primarily to the high frequencies of teacher-initiated procedural contacts with high expectation boys, and that this pattern, although still evident, was reduced somewhat in the second half of the data set.

The four remaining significant sex by expectancy groups interactions from Table 2 all fit into the same pattern and resemble similar interactions reported in several earlier studies (Brophy and Good, 1973). The four variables involved are total teacher-initiated private contacts (B14), percentage of correct answers followed by praise (E2), percentage of student-initiated work-related contacts in which the student received process feedback (G2), and the measure of the teacher's tendency to give the answer herself rather than call on someone else when the child failed to respond to the initial question (G4). All four of these interactions showed the high boys favored over the low boys and the low girls favored over the high girls. Also, in each case the variability among the boys was much greater than the variability among the girls.

These findings are consistent with earlier results suggesting that, because of their greater saliency, boys are more likely to be at the extremes of distributions of teacher-student interaction measures, while girls are more likely to be concentrated towards the means. As a result, high expectation boys generally show up as the most favored group and low expectation boys as the least favored group when a significant interaction is observed. The four significant interactions from Table 2 just described all fit this frequently observed pattern.

The same is true concerning the significant interaction for behavioral criticism (E11)that appears in Table 3 (center classes). This interaction reflects the fact that teachers were far more likely to criticize the low expectation boys for misbehavior than they were to criticize any of the other three groups. This interaction pattern has been observed in several other studies (Brophy and Good, 1973).

One other variable from Table 3 showed a significant expectancy groups by sex interaction, and also a significant expectancy groups by sex by trials interaction. This was the measure of the percentage of correct answers which were followed by praise from the teacher (E2). This time the interaction on the praise following correct answers variable was different from the more typical pattern seen in the regular classes. The basic reason for both the two-way and the three-way interaction was that the low expectation girls were particularly unlikely to be praised following correct answers, relative to the other three groups. However, this notable disadvantage relative to the other groups was confined mostly to the first half of the data set; by the second half the relative positions of the four groups had assumed the more typical pattern (hence the three-way interaction). It is not known whether this unusual interaction represents a reliable finding suggesting a difference in teacher praise behavior in center classes as opposed to regular classes. Although this is a possibility, the finding is perhaps best left uninterpreted pending replication.

In summary, most of the significant interactions between student sex and expectancy groups observed in the present study have been frequently observed in prior research (Brophy and Good, 1973). They reaffirm that the greater variability and saliency of boys interacts with expectancy group status so that the high expectancy boys generally have the most favorable patterns of interactions with the teachers and the low expectation boys the least favorable patterns, with the two groups of girls in between.

Summarizing the findings from the standard score analyses in Tables 2 and 3, it should be noted that although they brought out several interesting sex differences in teacher-student interaction patterns, they did not alter or add much to the conclusions concerning expectation effects drawn from the raw score analyses summarized in Table 1. They reconfirm the general conclusions that the Brophy and Good (1970a) findings were not replicated and that no evidence to support the polarization hypothesis was observed.

Data from Individual Classrooms

Despite the generally negative results described for the teachers as a group, a set of expectancy groups (2) by trials (2) analyses of variance in each of the 49 dependent classroom interaction measures was performed separately for each of the nine regular classrooms to

Table 4. Summary of Statistically Significant Expectancy Group Main Effects and Expectancy Groups by Trials Interactions with Each of the Nine Regular (Non-Center) Classrooms on Variables Related to Teachers' Communication of Performance Expectations.[1]

School	Lower Class					Middle Class				
Teacher	A		B		C	D		E		
Subject Matter[2]	LA	M	LA	M	M	M	M	LA	LA	
Ability Level of Class	Low	High	High	Low	Middle	Low	High	High	Low	Totals
Main Effects										
Favoring Highs	0	8	11	7	1	0	3	0	3	33
Favoring Lows	1	1	2	2	1	4	0	0	3	14
Interactions										
Favoring Highs	5	0	0	0	2	0	0	3	0	10
Favoring Lows	0	0	2	2	1	0	0	0	1	6

[1]These data are from expectancy groups (2) by trials (2) analyses of variance in the raw (unstandardized) frequency scores and percentage measures. Separate analyses were done for each of the nine classrooms. Data in this table were included only for variables related to teacher expectation effects (see text for explanation), and only when the main effect or interaction reached the .05 level of significance.

[2]LA = Language Arts; M = Mathematics

see if noteworthy patterns emerged for any of the individual teachers. The results of these analyses are shown in Table 4, which summarizes the significant ($p < .05$) main effects for expectancy groups and the significant expectancy groups by trials interactions on variables related to teacher communication of performance expectations. A few of the interaction measures were excluded from this table because they were not measures of teacher behavior (Cluster A) or because their implications regarding teachers' communication of performance expectations to students are ambiguous or uninterpretable. Variables in the latter category included reading turns (B2) and choice questions divided by the sum of process questions plus product plus choice (D2).

Removal of these five variables from the original set of 49 left a total of 44 possible variables which could have shown significant main effects or interactions. However, analyses of all 44 variables could not be performed for any of the nine classes involved. This was either because the relevant behavior was not observed in the

classroom (no data) or because all students in the class had scores of zero on the variable involved (no variance).

Inspection of Table 4 does reveal some interesting individual differences among the teachers. In particular, Teachers A and B (both at the lower class school) did show evidence of expectancy effects, while the other three teachers did not. This was most obvious with Teacher B, who favored the highs on 11 of 13 significant main effects in her high ability language arts class and also favored the highs on 7 of 9 significant main effects in her low ability math class. All 18 of the group differences on which this teacher favored highs over lows were from quantitative measures in Clusters B and C. Thus this teacher also fits the hypothesis that teacher expectation effects would be mediated through quantitative measures of interaction at the fifth grade level.

Inspection of the four measures on which she favored lows showed that she warned or criticized highs more frequently for misbehavior and more frequently failed to give feedback to highs following their response opportunities in the high ability language arts class, and that she initiated more work-related private interactions with lows and praised lows more frequently for good work in her low ability math class. Thus, three of the four differences favoring lows are on qualitative rather than quantitative measures.

The interaction data for Teacher B show that although she continued to favor the highs in the second half of the data set, she was slightly less favorable towards them in the second half. Thus, Teacher B's interaction data go against the polarization hypothesis. However, closer inspection of the variables involved shows that three of the four (student-initiated questions, self-reference questions, and opinion questions) are essentially trials effects rather than genuine interactions. These three categories of student response opportunity occurred very infrequently in the first half of the data set, and then all but disappeared in the second half. The few occasions on which behavior in these categories did occur involved the high expectation students almost exclusively; thus, their means on these three frequency scores dropped from very low to near zero between the first and the second data set, while the means for the low expectancy students were near zero in both sets. Thus, these three interactions reflect not so much a change in the teacher's behavior toward the high and low expectation students as the general trials effect described earlier: as the school year goes on, more time is spent on matters directly related to the curriculum and less on matters of personal interest to the student.

The fourth significant interaction for Teacher B(which occurred in her low ability math class) does appear to represent a significant group interaction. This is for variable B11, student-initiated private interactions (both work-related and procedural). This interaction denotes that as the school year progressed, the low expectancy students in this room, relative to the high expectancy students, increased

the frequency with which they initiated private interactions with the teacher. This is in direct contradiction to the polarization hypothesis, which predicts the opposite interaction.

In summary, the data for Teacher B show her to favor the highs over the lows in both a high ability language arts class and a low ability math class. They also show that this favoritism is mediated through quantitative rather than qualitative measures of interaction, as expected. However, the few significant expectancy groups by trials interactions go against the polarization hypothesis, which predicts that such favoritism of highs over lows will increase as the school year goes on. The data suggest that this teacher's favoritism decreased slightly, at least in one of her two classes.

The data for Teacher A also show favoritism of highs over lows, although the pattern is a little more complex than in the case of Teacher B. The main effects data show only one significant difference (the teacher more frequently failed to give feedback to highs than to lows) in the low ability language arts class, suggesting that the effects of expectancy groups in this class were minimal, and that the teacher did not favor one group over the other. However, interaction data for this same class show that as the year progressed the teacher did move toward beginning to favor the highs over the lows.

The five significant interactions all favored the highs, and all were on quantitative measures (direct questions, open questions, total response opportunities, total teacher-initiated private contacts, and total teacher-initiated procedural contacts). Furthermore, most of these differences were on variables that involved the teacher's initiation of a public or private contact with the student, suggesting that they do reflect a change in the teacher's behavior rather than simply an increased pressure from high ability students to seek response opportunities or initiate contacts with him. Thus, although the main effects show no clear favoritism within the time span included in the observation period, the interaction data suggest that the teacher was moving towards initiating more frequent contacts, both public and private, with high expectation students. This suggests that data taken in this classroom late in the spring would have shown a significant difference on several quantitative measures, favoring highs over lows.

The main effects data for Teacher A's other class (a high ability math class) at first suggest a clear-cut favoritism toward highs. Six of the eight differences favoring highs were on quantitative measures (open questions, call outs, total response opportunities, teacher-initiated procedural contacts, total student-initiated private contacts and call outs divided by the sum of direct questions plus open questions plus call outs). Most of these measures are on variables that are exclusively or primarily under the control of the students rather than the teacher, however, suggesting that the differences reflect a passive teacher reaction to behavioral differences

between the highs and lows in this classroom rather than a more proactive discrimination in favor of the highs. However, the highs were also favored on two qualitative measures. They received a significantly higher percentage of process questions, and they also received more process feedback in the work-related interactions than they initiated with the teacher. Thus, the teacher did appear to expect more from the highs and be more willing to work with them to help them understand the material. However, the single difference favoring lows also was on an important qualitative measure: the teacher stayed with lows following a wrong answer more frequently than he did with highs. Thus he also showed evidence of a willingness to work with lows, at least in these public response opportunity situations.

In summary, the main effects data for this classroom show the highs favored on eight of the nine significant differences. However, the pattern is such as to suggest that the teacher was only passively responding to differential student press rather than more proactively discriminating between the groups in communicating differential expectations. These group differences, as well as other aspects of interaction in this classroom, apparently were stable across the two halves of the data set, since no significant interactions were observed in this classroom.

The data for Teacher C show only two significant main effects and three significant interactions, and these are closely balanced between favoring highs and lows. Thus, expectancy groups were not an important factor in this class; Teacher C's teaching behavior apparently is not much affected by her performance expectations for students.

The data for Teacher D show weak and mixed results. In her low ability math class she had a slight but notable tendency to favor the lows. Although significant main effects were observed on only four measures, these were rather important measures of key teaching behavior. Teacher D asked the lows more direct questions, gave them more total response opportunities, and initiated more total private interactions with them in this class, and the lows responded by initiating more private interactions with the teacher than did the highs. These data suggest that Teacher D was making a concerted effort to reach the lows in this particular class. This pattern apparently was sustained across the two halves of the data set, since there were no significant interactions in this class.

In Teacher D's other class (a high ability math class) the three significant main effects observed all favored the highs. However, these differences were on somewhat less vital measures than the differences in the previous class. In this class, the highs received more open questions, more total response opportunities, and more process questions. The first two differences are more than likely due to persistence on the part of the highs in raising their hands and otherwise seeking public response opportunities. Thus, Teacher D's behavior in this high ability math class, like Teacher A's behavior

in his high ability math class, suggests a pattern of passive reaction to differential student press rather than proactive discrimination in favor of highs or against lows. This pattern apparently persisted across the data set, because again no significant interactions were observed in this class.

The data for Teacher E suggest that the expectancy groups were not an important factor in affecting teacher-student interaction in either of her two classrooms. There were no significant main effects in the high ability language arts class, although three significant interactions all favored the highs (opinion questions, teacher-initiated work-related interactions, and open questions divided by the sum of direct plus open plus call outs). These data suggest at best a slight tendency to interact more frequently with the highs as the school year progressed in this high ability language arts class.

Six significant main effects were noted in Teacher E's low ability language arts class, but they were split evenly between favoring highs and lows. In this class Teacher E initiated more procedural interactions and more total private interactions with lows, and she more frequently asked them a new question following a correct answer to an initial question. However, she also gave behavioral warnings and behavioral criticisms more frequently to lows, and she more frequently called on someone else to answer the question rather than give the answer herself when a low could not answer it. The contrast between her reaction to lows following correct answers and her reaction following their failures to answer correctly suggests that this teacher is more effective in dealing with lows when they are successful than she is when they are unsuccessful. In general, though, Teacher E did not favor either the highs or the lows in either of her classes. She seemed to be relatively unaffected by her expectations for students' performance. The one significant interaction in this class reflects the fact that Teacher E initiated relatively more procedural interactions with the lows as the year progressed. This goes against the polarization hypothesis, although procedural interactions are considered less important and less central to the communication of expectations than work-related interactions.

The data of Table 4 concerning the analyses within each of the nine regular classrooms can be summarized as follows. Teachers A and B, both working in the lower class school, tended to favor the highs in both of their classrooms. The great majority of the differences favoring highs in these four classrooms were on quantitative measures, as predicted. The interaction data relevant to the polarization hypothesis show negative results for the three classrooms in which clear-cut favoritism towards highs was observed, although the predicted polarization effect did occur in Teacher A's low ability language arts class, where the main effects data did not show favoritism towards highs.

The data for Teachers C, D, and E suggest that in general they were not much affected by the expectancy groups, although Teacher D did appear to be making a special effort to work with the low expectation students in her low ability math class. In general, the data from these three teachers bear out the broad conclusions already drawn: these teachers did not systematically favor the highs over the lows in their classroom interactions with students, nor did they show polarization effects as the school year progressed.

Stability over Time in Teacher-Student Interaction Patterns

Stability over time in patterns of teacher-student interaction was investigated by correlating (within each classroom) each of the 49 interaction measures from the first half of the data set with the same measure from the second half of the data set. These data are presented in Table 5, separately for the five regular classrooms in the lower class school, for the four regular classrooms in the middle class school, and for the four center classes in the middle class school. The data include a percentage breakdown of the total distributions of the five possible outcomes from the stability coefficient analyses, as well as percentage breakdowns within the three possibilities when correlation coefficients were run. The latter possibilities included significant ($p < .05$) positive correlations, nonsignificant positive correlations, and nonsignificant negative correlations (no significant negative correlations were observed).

In addition to these three possible outcomes when correlation coefficients were computed, there were two situations in which correlation coefficients could not be computed. The first of these occurred when behavior relevant to a given interaction measure was observed in each half of the data in a given classroom, but all students received a score of zero on this behavior. In a sense, this means that the correlation between the two halves of the data set was perfect, but it is more statistically correct to say that correlation coefficients could not be computed due to lack of variance in one or both of the data sets. Thus in Table 5, as well as in Tables 6 and 7 to follow, the percentage figures given in the rows entitled "No Correlations Due to Lack of Variance" refer to the situation in which all subjects within a classroom were scored zero on a given variable in one or both of the data sets, so that correlation coefficients could not be computed due to lack of variance.

The percentage figures in Tables 5, 6, and 7 given in the rows labeled "No Data; Behavior Did Not Occur" refer to situations where the behavior corresponding to a given variable was not observed at all or was observed for only one subject, so that a correlation coefficient could not be computed. This situation occurred frequently, especially in center classes, for variables dealing with teachers' reactions to students' responses. For example, a stability coefficient could not be computed for the variable "Praise following part-correct answer" (E3) if no student in the class was ever coded for a part-correct answer, or if only one student was coded for a part-correct answer.

Table 5. Distributions of Stability Coefficients Reflecting the Correlations between First Half and Second Half Percentage Scores and Frequency Measures within Each Classroom.[1]

School	Lower	Middle	
Type of Class	Regular	Regular	Center
Number of Classes	5	4	4
Total Distributions			
Significant Positive Correlations[2]	20%	15%	12%
Non-Significant Positive Correlations	29%	21%	9%
Non-Significant Negative Correlations	25%	13%	14%
No Correlations Due to Lack of Variance[3]	18%	28%	31%
No Data; Behavior Did Not Occur[3]	8%	22%	35%
Distributions of Computed Stability Coefficients			
Significant Positive Correlations[2]	27%	31%	34%
Non-Significant Positive Correlations	39%	43%	25%
Non-Significant Negative Correlations	34%	26%	41%

[1]The stability coefficients summarized in this table are Pearson r's between students' first half and second half scores, computed separately for each class, using the raw (unstandardized) frequency scores and percentage measures.

[2]$p < .05$.

[3]See text for explanation.

Although these last two situations (lack of variance in the scores observed or failure to observe the relevant behavior) represent a form of consistency in teacher and student behavior, they do not allow computation of statistics such as correlation coefficients. They were included in Tables 5, 6, and 7, however, to help give perspective to the three kinds of correlation coefficients which were observed.

The data in Table 5 show that there is moderate stability in the interaction measures between the first and second halves of the data set, although some differences by type of classroom are apparent. First, comparing only the regular classrooms, it is clear that stability was higher in the middle class school than the lower class school. Among those correlation coefficients which were computed, 31% of those at the middle class school were significant positive correlations, 43% were nonsignificant but positive, and only 26% were

negative. The corresponding figures for the lower class school were 27%, 39%, and 34%. Thus the middle class school had higher percentages of both types of positive correlations and a lower percentage of nonsignificant negative correlations.

Inspection of the total distribution data at the top of Table 5 suggests that the school difference may be primarily due to differences in the number of behavior categories used at the two schools rather than to a greater stability at the middle class school on the common categories. Note that 74% of the possible correlation coefficients were computed at the lower class school, while only 49% of the possible coefficients were computed in the regular classes at the middle class school. This means that many more correlations were computed from the data from the lower class school on variables which had relatively low frequencies of occurrence in the classroom. Such variables are less likely to show stability than are variables which are more frequently observed. Thus, the suggestion that stability was higher at the middle class school in the regular classes must be taken with caution, since it may be an artifact of the number of behavioral categories used at the two schools rather than a true difference in the stability of teacher or student behavior.

Similar comments apply to the difference within the middle class school between the regular and the center classes. Although 49% of the possible correlation coefficients were computed in the regular classes, only 35% of the possible correlation coefficients were computed for the center classes. Thus, more of the possible behavior categories were used in the regular classes. Nevertheless, stability was greater in the regular than in the center classes. Among the coefficients computed, 31% of those in the regular classes were significant and positive, 43% were positive but not significant, and only 26% were nonsignificant and negative. The corresponding figures for the center classes were 34%, 25%, and 41%. Thus, although the center classes did have a slightly higher percentage of significant positive correlations, they also had a much higher percentage of negative correlations. Thus, stability across the two halves of the data set was higher for the regular classes than for the center classes, even though stability coefficients were computed for a greater percentage of the possible total in the regular classes. This difference could be due in part to differences in the rates of the interaction in the two types of classrooms. Not only was a smaller percentage of possible behavioral categories used in the center classrooms; fewer instances of these behaviors were observed in the centers than were observed in parallel behavioral categories in the regular classes. This was because teacher-student interaction was much less frequent in the center classes than in the regular classes. Thus, interaction rates may have had some effect on the differences in stability percentages in the two types of classes.

It is also likely that the differences in teacher roles in the two types of classes affected these stability percentages. To the extent that teachers had a stable style which tended to structure

the classes they taught, stability would be higher in regular classes than in center classes. This is because the teachers carry on much more structured teaching in regular classes, and the events occurring in these classes are primarily planned and executed by them. In contrast, the teacher in the center classroom is primarily a proctor or supervisor whose role is confined mostly to keeping order and responding to individual needs. Thus, in this setting the teacher is much more reactive and much less proactive than in the regular classroom, and any stability in the data which is due to regularities in teaching style would not show up in these center classes.

In summary, the data of Table 5 suggest that stability in teacher-student interaction patterns was greater at the lower class school than at the middle class school, and that within the middle class school it was greater in the regular classrooms than in the center classes. Both of these conclusions must be stated with caution, however, in view of possible artifacts which may be affecting them.

Cross-Class Correlations

The data in Table 5 concerned the stability coefficients within each classroom between the first and second halves of the data set. In contrast, the data in Table 6 reflect the degree of stability across classrooms within either half of the data set. The percentage data in Table 6 come from correlation coefficients in which a given interaction measure representing a teacher-student interaction pattern in one class was correlated with the corresponding measure involving the same student in a different class. The different class usually involved the same student interacting with a different teacher, although in a few cases the same teacher was involved. All of these cross-class correlations were computed within the two halves of the data set. Thus, a particular type of teacher-student interaction pattern within one classroom in the first half of the data set was correlated against the same pattern in another classroom in the first half of the data set. This same correlation was then repeated for the two classrooms in the second half of the data set.

Table 6 presents percentage figures summarizing the results of these cross-class correlations. The data for the lower class school are from the five regular classes combined. The data from the middle class school are broken into three sets, involving correlations between two regular classes, correlations between two center classes, and correlations between one center class and one regular class. Two different teachers were involved in each set of correlations except for the latter category; half of the correlations between a regular class and a center class involved the same teacher in each instance. These will be discussed below.

Table 6. Distributions of Stability Coefficients Reflecting the Correlations between Students' Frequency Scores and Percentage Measures from Two Different Classrooms. [1]

School	Lower	Middle		
Type of Class	Regular Only	Regular Only	Center Only	One of Each
Number of Pairs of Classes Correlated	3	2	2	8
Total Distributions				
Significant Positive Correlations[2]	9%	8%	6%	9%
Non-Significant Positive Correlations	28%	18%	15%	17%
Non-Significant Negative Correlations	33%	19%	8%	10%
No Correlations Due to Lack of Variance[3]	21%	33%	33%	36%
No Data: Behavior Did Not Occur[3]	10%	23%	38%	29%
Distributions of Computed Stability Coefficients				
Significant Positive Correlations[2]	13%	17%	21%	25%
Non-Significant Positive Correlations	40%	38%	51%	47%
Non-Significant Negative Correlations	48%	44%	28%	28%

[1]The stability coefficients summarized in this table are Pearson r's between students' raw (unstandardized) frequency scores and percentage measures from one class with their corresponding measures from another class, computed within either the first half or the second half of the data set. Data from the two halves were then combined to compute the percentage shown in the table.

[2]$p \leq .05$.

[3]See text for explanation.

Two major findings are notable in the data of Table 6. First, comparing it with the data of Table 5, it is clear that, as might have been expected, there is less stability across different classrooms within the same time period than there is across time periods within

the same classroom. Thus, a given student's patterns of interaction with his teacher in a given class are more stable over time than are his interactions within a shorter time period with different teachers in different classrooms.

Secondly, comparing the data for regular classes only with the data involving centers in Table 6, it is clear that stability was lower for correlations involving two different regular classes than for correlations involving two centers or involving one center and one regular class. This was somewhat unexpected, especially in view of the data of Table 5 showing generally higher stability within regular classes than centers at the middle class school. Most likely, the relatively lower stability percentages in the data in Table 6 for regular classes result from subject matter differences. In each case, correlations between two regular classes involved correlating a language arts class with a mathematics class. Thus, although the same student was involved in a given pair of interaction scores, both the teacher and the subject matter were different. Of these factors, the differences in the kinds of activities that were included in the structured mathematics and language arts classes were probably most responsible for these lower stability coefficients for cross-class correlations involving these classes.

Subject matter also differed for the cross-class correlations involving center classes (Column 3 of Table 6), but the kinds of activities that went on the mathematics and language arts centers did not differ so much from each other as the kinds of activities that went on during the language arts and mathematics structured classes. That is, regardless of whether a center was a language arts center or a math center, the students still primarily worked at individual assignments and the teacher did not attempt to do structured group teaching. Thus, the center settings were quite comparable despite the differences in subject matter title.

The eight pairs of classes involved in the correlations between a regular class and a center class were divided into two subtypes in order to investigate the effects of same vs. different teacher and same vs. different subject matter. Four of these classes involved the same teacher but different subject matter (as when a given teacher had the same group of students for a structured math class and a language arts center class, or vice versa). The other four classes involved different teachers but the same subject matter (the students always had one of the two teachers for their structured class in a given subject and the other teacher for their center class in that same subject).

Comparisons of these two subsets of correlations involving regular classes and centers showed almost exactly equivalent percentage distributions. In each case, 36% of the possible correlations that could have been computed were computed, while 64% were not computed due to lack of variance or failure of the behavior involved to occur. Furthermore, when the figures for the three types of coefficients computed are broken down, the results show very close compara-

bility. For the four pairs of classes involving the same teacher but different subject matter, 25% of the correlations were positive and significant, 53% positive but not significant, and 22% negative and not significant. The corresponding figures for the four classes involving different teachers with the same subject matter were 25%, 42%, and 33%, respectively. Thus, each type of situation produced the same percent (25%) of positive significant correlations. The figures for the nonsignificant correlations show a slight advantage favoring the four classes in which the teachers were the same over the four classes involving different teachers but the same subject matter. However, these differences seem relatively minor in the context of the larger comparability of the two sets of data.

The data were also analyzed to see if high ability vs. low ability students made a difference. These comparisons all showed that regardless of teacher, school, or type of class (regular vs. center), student ability made no difference on the relative percentages of significant positive, nonsignificant positive, and nonsignificant negative stability coefficients. Thus, student ability level appears to be unrelated to stability across time within class or stability across classes within the same time period in teacher-student interaction patterns.

In summary, the most important finding from Table 6 was that these correlations were generally lower than the correlations in Table 5, showing that stability in teacher-student interaction patterns across two different classes is somewhat lower than stability over time in teacher-student interaction patterns within the same classroom. Also, stability coefficients were generally lower when only regular classes were involved than when centers were involved. The apparent reason for this was that the correlations involving only regular classes all were between language arts classes and mathematics classes, and the differences between the kinds of activities that go on in structured classes in these two subject matter areas apparently reduced the stability of teacher-student interaction patterns in them as compared to the other kinds of class comparisons included in Table 6.

Stability of the Seven Clusters of Interaction Variables

Table 7 is based on the same set of cross-class correlations that formed the basis for Table 6. However, this time the coefficients for different types of classes have been combined and then retabulated separately for each of the seven clusters of interaction variables, so that comparisons among these seven clusters on the degree of stability across classes could be facilitated. Several points of interest are notable in Table 7.

First, the total distribution data show that Clusters B and C, which basically involved the frequency measures rather than the qualitative percentage scores, were observed in all classrooms. Also,

Table 7. Distribution of Stability Coefficients Reflecting the Correlations between Students' Frequency Scores and Percentage Measures from Two Different Classes, Tabulated Separately for Each of the Seven Clusters of Interaction Variables.[1,2]

	Interaction Variable Clusters						
Total Distributions	A	B	C	D	E	F	G
Significant Positive Correlations[3]	3%	18%	15%	3%	4%	1%	1%
Non-Significant Positive Correlations	27%	36%	36%	20%	4%	2%	10%
Non-Significant Negative Correlations	12%	19%	29%	23%	9%	8%	17%
No Correlations Due to Lack of Variance[4]	21%	28%	20%	53%	48%	30%	47%
No Data; Behavior Did Not Occur[4]	37%	0%	0%	0%	35%	60%	25%
Distributions of Computed Stability Coefficients							
Significant Positive Correlations[3]	7%	25%	18%	7%	24%	4%	2%
Non-Significant Positive Correlations	64%	50%	45%	43%	24%	20%	36%
Non-Significant Negative Correlations	29%	26%	37%	50%	53%	76%	62%

[1]The stability coefficients summarized in this table are the same Pearson r's tabulated in Table 6, except that data from the different types of classes (as well as the two halves of the data set) were combined before percentages were computed.

[2]Cluster A = student performance indicators; Cluster B = frequencies of each type of teacher-student interaction; Cluster C = teacher vs. student initiation of contacts; Cluster D = type of teacher questions; Cluster E = teacher praise and criticism; Cluster F = teacher persistence in eliciting responses; Cluster G = level of feedback given to students

[3]$p < .05$.

[4]See text for explanation.

72% of the possible correlations for Cluster B and 80% of the possible correlations for Cluster C were actually computed because the behavioral events for these variables not only were observed but occurred with sufficient variability to allow computation of correlation coefficients. These percentages are much higher than the corresponding percentages for the other five clusters of variables. Some clusters, particularly Clusters E and F, occurred relatively infrequently.

Cluster E concerned teacher praise and criticism. In this sample teacher praise was relatively infrequent, and teacher criticism was

very rarely coded, except for criticism for misbehavior. Criticisms for wrong answers or for failure to respond were almost nonexistent. The low frequencies in Cluster F dealing with the quality of teacher feedback to student responses occurred because certain types of student responses were themselves infrequent (part-correct answers, "don't know" responses, and no responses). Because of this, no data for teacher feedback in these situations could be coded, either. The unusual figures for Cluster D appear primarily because process questions were infrequent, so that correlations for the variable process questions divided by the total of process plus product plus choice questions (D1) often could not be computed because all students in the class had a score of zero on this variable.

Inspection of the distributions of those stability coefficients which were computed shows that Clusters B and C again show a difference from the other clusters. These two clusters had many more positive significant correlations and many fewer negative correlations than the other clusters. Cluster A, dealing with student responses, also had few negative correlations but also had relatively few significant positive correlations.

More generally, the data for the distribution of coefficients actually computed in Table 7 bring out a point that must be taken into account in viewing all of the stability data of this study: stability tends to be positively associated with the frequency with which a given behavioral category was observed in the data. In many cases the correlation computations for a given pair of score distributions were based upon a low average of occurence per student of the behavior involved and/or a low number of students for whom the behavior was observed. This situation makes for low stability scores and sometimes requires very high positive correlation in order for the coefficient to reach statistical significance.

As a result, the stability data given in Tables 5, 6, and 7 give a somewhat lower impression of the general stability in the data than was actually the case, because all coefficients actually computed were included in computing the percantage data shown in these tables, even if the behavior involved had a very low frequency of observation and even if data in a given class were available on only three or four students. If minimum cutoff points regarding frequency of occurrence of the behavior and/or number of students for whom data were available had been established, the stability coefficients for all variables might have been similar to those for Clusters B and C in Table 7. Thus, roughly about 20% of the correlatinns would have been significant and positive, about 50% positive but not significant, and about 30% negative (and not significant). Although this represents a moderate degree of stability, such data are not nearly as impressive as the stability data for high inference ratings or other measures of classroom interaction which are based upon global inferences or impressions rather than coded observations of discrete interactions. This is consistent with previous findings regarding observations of discrete classroom behaviors (Rosenshine and Limbacher, 1972).

DISCUSSION

The main purposes of this study were to see if the Brophy and Good (1970a) findings from the first grade level regarding teachers' communication of performance expectations would be replicated at the fifth grade level, and to test the hypothesis that if such findings were replicated the class would show polarization over time as the school year progressed. The findings regarding both of these questions were almost completely negative. There was little evidence that teachers' expectations for student performance made much difference in their treatment of different students in the same classroom, and few of the Brophy and Good (1970a) findings were replicated in the present study. On account of this, the polarization hypothesis could not even be tested, since it assumes expectation effects and cannot be tested where no expectation effects exist.

A secondary hypothesis investigated in the data was that expectation effects at the fifth grade level would be more likely to be mediated through quantitative than qualitative measures of teacher-student interaction. This hypothesis did receive some support in that the predicted significant expectancy group differences which did appear were mostly on quantitative rather than qualitative measures of teacher-student interaction, and the significant reversals of previous findings were on qualitative measures. However, this support for the secondary hypothesis must be viewed within the larger context of generally weak and negative findings for expectations.

The findings of the present study are remarkably parallel to findings from a followup first grade study (Brophy and Good, 1973). They demonstrate once again that expectation effects are not necessary or universal, even when teachers' naturalistically formed expectations are used as the basis of the investigation (as opposed to expectations induced in the teachers through some kind of experimental manipulation).

The present data are also consistent with a number of observations made by Brophy and Good (1973) concerning the conditions under which expectation effects are or are not likely to be observed. Their review of the literature shows that expectation effects are more often observed when the contacts between the teachers and students are brief rather than extended over a period of months, and more likely to be observed when the data are collected toward the end of the school year than toward the beginning of the year. The present data were collected primarily in the first half of the school year, and the period of observation extended over several months, thus reducing the likelihood of observing expectation effects.

However, analyses of the individual teachers' data showed clear expectation effects for one teacher and the strong suggestion of such effects for another. This is consistent with data from several sources

suggesting that individual differences among teachers are important in determining whether or not expectation effects are observed (or, to phrase it differently, whether or not teachers allow their expectations to affect the way they treat students). On the basis of data from several studies, Brophy and Good (1973) argue that the presence of expectation effects is an indication of relatively poor teaching, and that the more competent a teacher is the less likely expectation effects are to occur in his classroom. Internal analyses of the present data lend support to this interpretation.

Compared to the three teachers who showed no evidence of expectation effects, the two teachers who did show such evidence appeared to be less competent teachers. For example, their students typically created fewer work-related contacts with them, and a much greater percentage of the private teacher-student contacts were procedural contacts rather than work-related contacts, suggesting less organization and less concern about achievement in these two teachers. Also, behavioral warnings and criticisms were much more frequent in their classrooms than in the others'; suggesting that they had more difficulty maintaining classroom order and/or that they were criticism- rather than praise-oriented in their approach to student motivation. They also had notably lower percentages of correct answers over total answers, suggesting that they were not propoerly adapting the material to the abilities of their students and/or that their students hesitated to respond unless they knew the right answer. These teachers also more frequently failed to give feedback following student responses in comparison to other teachers. Thus, the internal evidence suggests that the two teachers who did show a tendency to communicate expectations to their students were less competent teachers than the other three teachers on a variety of measures of teacher-student interaction. These findings tie in with similar findings reported by Brophy and Good (1973) from several other studies.

The most probable explanation for the relationship between expectation effects and teacher competence is that the exceptionally competent or talented teacher has a broad repertoire of skills to bring to bear in diagnosing and remediating learning difficulties, and that in his moment-to-moment interactions with students he remains problem-centered and draws on this repertoire to overcome any difficulties encountered. In other words, this type of teacher probably remains problem-centered in the face of difficulties, shifting to a different strategy when the one he is using doesn't work. In contrast, the less competent teacher has a more limited repertoire of diagnosis and remediation skills, so that he is less likely to be able to remain problem-centered in the face of persistent learning difficulties with certain students. This teacher, if he has tried everything he knows and still has not succeeded, will be more prone to giving up on the student and beginning to rationalize his failure or seek excuses for it. Once this process of psychologically giving up on a student has begun, the potential for expectation effects increases. Once expectation effects begin to actually occur, the vicious circle described earlier gets set into motion so that it becomes self-reinforcing.

The polarization hypothesis derived from Brophy and Good's (1970a) model for expectation effects still remains essentially untested, despite two attempts to test it. In both cases it could not be tested because an adequate number of teachers did not show expectation effects to allow a clear test. The findings in the present study were particularly negative, however, in that analyses of the groups by trials interactions in the classrooms of the two teachers who did show expectancy effects provided little support for the polarization hypothesis. Thus, to date there still is no evidence to support the hypothesis that highs and lows become more different from each other as the school year goes on in classrooms where the teacher's teaching is affected by his expectations for students.

Continued failure to find support for the polarization hypothesis may force a revision of the Brophy and Good (1970a) model for expectation effects. This step is not essential to the model since differential teacher treatment of different students (Step 2 of the model) could by itself affect student achievement by affecting student opportunity to learn. This could occur even if such differential teacher treatment did not lead to complementary student response and therefore polarization of the class over time. Thus, it is possible, for example, for a teacher to favor highs by having many more interactions with them and in general being more positive with them than with lows, but for the teacher's students to not allow this differential treatment to affect the way that they respond to the teacher. If this were to happen, the relative advantage of highs over lows would remain constant over the course of the school year rather than increase over time as the polarization hypothesis suggests. In any case, however, it is likely that the preferential treatment on the part of the teacher would cause the highs to achieve at or near their potential, while it would tend to cause the lows to achieve at a level somewhat below their potential. Thus, differential teacher expectations could become self-fulfilling even if students did not respond reciprocally to differential teacher behavior.

The special circumstances of this study allowed an investigation of the stability of teacher-student interaction patterns when students were in classrooms taught by two different teachers or, in a few cases, in different classrooms taught by the same teachers. In general, the stability data were not very impressive, although moderate stability was shown for certain categories of interaction. The low frequencies of observation of behavior relevant to certain categories suggest that the stability coefficients may be lower than would be the case if only high frequency categories had been included in the analyses or if more data had been collected. Nevertheless, they show that, as had been pointed out elsewhere (Rosenshine and Limbacher, 1972), classroom interaction data based on coded observations of discrete behaviors are less reliable (although not necessarily less valid) than high inference ratings or other data based on observers' global judgments.

The degree to which the stability data from this study are representative is unknown, since they are unique data. In any case, they suggest that student ability level does not affect the stability of classroom interaction measures, and also that correlations between measures taken in different subject matter classes taught by the same teacher tend to be only very slightly higher than correlations taken in classes involving the same subject matter taught by two different teachers. Thus, correlations between sets of data from two different classes will vary according to whether the same teacher or two different teachers are involved, whether the subject matter is the same or different, and whether the class is a regular class or a center or some other unusual type of class. Further conclusions from these stability data should be reserved until replication studies or comparison data are available.

REFERENCES

Beez, W. Influence of biased psychological reports on teacher behavior and pupil performance. Proceedings of the 76th Annual Convention of the American Psychological Association, 1968, 605-606.

Brophy, J. and Good, T. Teachers' communication of differential expectations for children's classroom performance: some behavioral data. Journal of Educational Psychology, 1970a, 61, 365-374.

Brophy, J. and Good, T. Brophy-Good System (Teacher-Child Dyadic Interaction). In A. Simon and E. Boyer (Eds.), Mirrors for behavior: an anthology of observation instruments continued, 1970 supplement. Volume A. Philadelphia: Research for Better Schools, Inc., 1970b.

Brophy, J. and Good, T. Teacher expectations: beyond the Pygmalion controversy. Phi Delta Kappan, 1972, in press.

Brophy, J. and Good, T. Classroom interaction: the influence of teacher perception. New York: Holt, Rinehart & Winston, 1973, in press.

Doyle, W., Hancock, G. and Kifer, E. Teachers' perceptions: do they make a difference? Paper presented at the Annual Meeting of the American Educational Research Association, 1971.

Elashoff, J. and Snow, R. Pygmalion reconsidered. Belmont, California: Charles A Jones, 1971.

Horn, E. Distribution of opportunity for participation among the various pupils in the classroom recitations. New York: Teachers College, Columbia University, 1914.

Jackson, P. and Lahaderne, H. Inequalities of teacher-pupil contacts. Psychology in the Schools, 1967, 4, 204-211.

Medinnus, G. and Unruh, R. Teacher expectations and verbal communication. Paper presented at the Annual Meeting of the Western Psychological Association, 1971.

Meichenbaum, D., Bowers, K., and Ross, R. A behavioral analysis of teacher expectancy effect. Journal of Personality and Social Psychology, 1969, 13, 306-316.

Palardy, J. What teachers believe, what children achieve. Elementary School Journal, 1969, 69, 370-374.

Rist, R. Student social class and teacher expectations: the self-fulfilling prophecy in ghetto education. Harvard Educational Review, 1970, 40, 411-451.

Rosenshine, B. and Limbacher, P. Relationship of high inference and low inference observation measures. Paper presented at the Annual Meeting of the American Educational Research Association, 1972.

Rosenthal, R. and Jacobson, L. Pygmalion in the classroom: teacher expectation and pupils' intellectual development. New York: Holt, Rinehart & Winston, 1968.

Rothbart, M., Dalfen, S., and Barrett, R. Effects of teacher expectancy on student-teacher interaction. Journal of Educational Psychology, 1971, 62, 49-54.

Rowe, M. Science, silence, and sanctions. Science and Children, 1969, 6, 11-13.

Rubovits, P. and Maehr, M. Pygmalion analyzed: toward an explanation of the Rosenthal-Jacobson findings. Journal of Personality and Social Psychology, 1971, 19, 197-203.

Tuckman, B. and Bierman, M. Beyond Pygmalion: Galatea in the schools. Paper presented at the Annual Meeting of the American Educational Research Association, 1971.

APPENDIX A.
TEACHERS' RANKING FORM

Teacher_____ Room_____ School_____
Subject_____ Time__:__ to __:__
Group Ability Level_____

Please list your students in order of achievement. The best student should receive rank "1," the next best should receive rank "2," etc.

(Top Student)

1. _____
2. _____
3. _____
4. _____
5. _____
6. _____
7. _____
8. _____
9. _____
10. _____
11. _____
12. _____
13. _____
14. _____
15. _____
16. _____
17. _____
18. _____
19. _____
20. _____
21. _____
22. _____
23. _____
24. _____
25. _____
26. _____
27. _____
28. _____
29. _____
30. _____
31. _____
32. _____
33. _____
34. _____
35. _____
36. _____
37. _____
38. _____
39. _____
40. _____

(Bottom Student)